

Petition for *Inter Partes* Review of  
U.S. Patent No. 8,015,006 B2

UNITED STATES PATENT AND TRADEMARK OFFICE

---

BEFORE THE PATENT TRIAL AND APPEAL BOARD

---

GOOGLE LLC,  
Petitioner

v.

DIALECT, LLC,  
Patent Owner

---

Case IPR2024-00748  
U.S. Patent No. 8,015,006 B2  
Issue Date: September 6, 2011

Title: SYSTEMS AND METHODS FOR  
PROCESSING NATURAL LANGUAGE SPEECH UTTERANCES  
WITH CONTEXT-SPECIFIC DOMAIN AGENTS

**PETITION FOR *INTER PARTES* REVIEW  
OF U.S. PATENT NO. 8,015,006 B2**

## Table of Contents

|                                                                                                    | Page |
|----------------------------------------------------------------------------------------------------|------|
| I. MANDATORY NOTICES UNDER §42.8(A)(1) .....                                                       | 15   |
| A. Real Party-In-Interest under §42.8(b)(1).....                                                   | 15   |
| B. Related Matters under §42.8(b)(2) .....                                                         | 15   |
| C. Lead and Back-Up Counsel under §42.8(b)(3) .....                                                | 16   |
| D. Service Information .....                                                                       | 17   |
| II. FEE PAYMENT .....                                                                              | 17   |
| III. REQUIREMENTS UNDER §§ 42.104 AND 42.108 AND<br>CONSIDERATIONS UNDER §§ 314(A) AND 325(D)..... | 17   |
| A. Grounds for Standing .....                                                                      | 17   |
| B. Identification of Challenge and Statement of Precise Relief<br>Requested .....                  | 17   |
| C. Considerations Under §§ 314(a) and 325(d).....                                                  | 18   |
| IV. OVERVIEW OF THE CHALLENGED PATENT.....                                                         | 19   |
| A. Person of Ordinary Skill in the Art (POSITA) .....                                              | 19   |
| B. Specification Overview .....                                                                    | 19   |
| V. CLAIM CONSTRUCTION .....                                                                        | 21   |
| VI. THE CHALLENGED CLAIMS ARE UNPATENTABLE.....                                                    | 21   |
| A. Overview of Grounds .....                                                                       | 21   |
| B. Ground 1: Claims 1-3 Are Obvious Over Coffman in View of<br>Kanevsky, Kneser and Zadrozny. .... | 22   |
| 1. Claim 1: “A method...” .....                                                                    | 22   |
| (a) “receiving...” (1[a]).....                                                                     | 23   |
| (b) “recognizing...” (1[b]) .....                                                                  | 25   |
| (i) “dynamically updating...” (1[b][i]) .....                                                      | 31   |
| (ii) “determining...” (1[b][ii]).....                                                              | 33   |
| (iii) “associating...” (1[b][iii]) .....                                                           | 35   |

# **Table of Contents** (continued)

|                                                                                                                                     | <b>Page</b> |
|-------------------------------------------------------------------------------------------------------------------------------------|-------------|
| (c) “parsing...” (1[c]).....                                                                                                        | 39          |
| (d) “formulating...” (1[d]) .....                                                                                                   | 42          |
| (e) “processing...” (1[e]) .....                                                                                                    | 49          |
| (f) “presenting...” (1[f]) .....                                                                                                    | 51          |
| 2. Claim 2.....                                                                                                                     | 52          |
| (a) “matching...” (2[a]).....                                                                                                       | 52          |
| (b) “scoring...” (2[b]).....                                                                                                        | 53          |
| (c) “selecting...” (2[c]) .....                                                                                                     | 55          |
| 3. Claim 3.....                                                                                                                     | 55          |
| (a) “requesting...” (3[a]).....                                                                                                     | 55          |
| (b) “determining...” (3[b]) .....                                                                                                   | 56          |
| C. Ground 2: Claim 4 Is Obvious Over Ground 1 Prior Art in<br>Further View of Maes ’101 and Franceschi.....                         | 56          |
| 1. Claim 4.....                                                                                                                     | 56          |
| (a) “receiving...” (4[a]).....                                                                                                      | 56          |
| (b) “recognizing...” (4[b]) .....                                                                                                   | 57          |
| (i) “requesting...” (4[b][i]).....                                                                                                  | 57          |
| (ii) “updating...” (4[b][ii]) .....                                                                                                 | 61          |
| (c) “parsing...” (4[c]).....                                                                                                        | 62          |
| (d) “formulating...” (4[d]) .....                                                                                                   | 62          |
| (e) “processing...” (4[e]) .....                                                                                                    | 62          |
| (f) “presenting...” (4[f]) .....                                                                                                    | 62          |
| D. Ground 3: Claim 5 Is Obvious Over Ground 1 Prior Art<br>(Coffman, Kanevsky, Kneser and Zadrozny) in Further View of<br>Chai..... | 63          |
| 1. Claim 5.....                                                                                                                     | 63          |
| (a) “receiving...” (5[a]).....                                                                                                      | 63          |

# **Table of Contents** (continued)

|                                                                                                                                                           | <b>Page</b> |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------|-------------|
| (b) “recognizing...” (5[b]) .....                                                                                                                         | 63          |
| (c) “parsing...” (5[c]).....                                                                                                                              | 63          |
| (d) 63                                                                                                                                                    |             |
| (e) “formulating...” (5[d]) .....                                                                                                                         | 63          |
| (i) “determining...” (5[d][i]).....                                                                                                                       | 64          |
| (ii) “extracting...” (5[d][ii]) .....                                                                                                                     | 69          |
| (iii) “inferring...” (5[d][iii]) .....                                                                                                                    | 71          |
| (iv) “transforming...” (5[d][iv]).....                                                                                                                    | 72          |
| (f) “processing...” (5[e]) .....                                                                                                                          | 74          |
| (g) “presenting...” (5[f]) .....                                                                                                                          | 74          |
| E. Ground 4: Ground 4: Claims 6 and 9 Are Obvious Over<br>Ground 1 Prior Art (Coffman, Kanevsky, Kneser and<br>Zadrozny) in Further View of Redfern. .... | 74          |
| 1. Claim 6.....                                                                                                                                           | 74          |
| (a) “receiving...” (6[a]).....                                                                                                                            | 74          |
| (b) “parsing...” (6[b]).....                                                                                                                              | 78          |
| (c) “formulating...” (6[c]).....                                                                                                                          | 78          |
| (d) “processing...” (6[d]) .....                                                                                                                          | 79          |
| (e) “directing...” (6[e]) .....                                                                                                                           | 79          |
| (f) “querying...” (6[f]).....                                                                                                                             | 80          |
| (g) “submitting...” (6[g]) .....                                                                                                                          | 81          |
| (h) “asynchronously receiving...” (6[h]) .....                                                                                                            | 83          |
| (i) “scoring...” (6[i]).....                                                                                                                              | 83          |
| (j) “determining...” (6[j]).....                                                                                                                          | 84          |
| (k) “applying...” (6[k]) .....                                                                                                                            | 86          |
| (l) “evaluating...” (6[l]).....                                                                                                                           | 87          |
| (m) “selecting...” (6[m]) .....                                                                                                                           | 88          |

**Table of Contents**  
(continued)

|                                     | <b>Page</b> |
|-------------------------------------|-------------|
| (n)   “presenting...” (6[n]).....   | 89          |
| 2.    Claim 9.....                  | 90          |
| (a)   “determining...” (9[a]) ..... | 90          |
| (b)   “requesting...” (9[b]).....   | 91          |
| VII.   CONCLUSION.....              | 92          |
| CERTIFICATE OF SERVICE .....        | 94          |

**List of Exhibits**

| <b>Exhibit No.</b> | <b>Description of Document</b>                                                                                                                                                      |
|--------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>1001</b>        | U.S. Patent No. 8,015,006 B2 to Robert Kennewick et al. (filed May 30, 2008, issued September 6, 2011) (“ <b>006</b> ” or “ <b>006 patent</b> ”)                                    |
| <b>1002</b>        | Declaration of Christopher M. Schmandt (“ <b>Schmandt</b> ”)                                                                                                                        |
| <b>1003</b>        | International Patent Application Publication No. WO 00/20962 to Daniel Coffman, Stephane H. Maes et al. (filed October 1, 1999, published April 13, 2000) (“ <b>Coffman</b> ”)      |
| <b>1004</b>        | U.S. Patent No. 5,897,616 to Dimitri Kanevsky and Stephane Herman Maes (filed June 11, 1997, issued April 27, 1999) (“ <b>Kanevsky</b> ”)                                           |
| <b>1005</b>        | U.S. Patent No. 6,157,912 to Reinhard Kneser et al. (filed March 2, 1998, issued December 5, 2000) (“ <b>Kneser</b> ”)                                                              |
| <b>1006</b>        | Wlodek Zadrozny et al., “Natural Language Understanding with a Grammar of Constructions”, Proc. 15th Int’l Conf. on Computational Linguistics (Vol. 2) (1994) (“ <b>Zadrozny</b> ”) |
| <b>1007</b>        | International Patent Application Publication No. WO 00/21075 to Stephane H. Maes et al. (filed October 1, 1999, published April 13, 2000) (“ <b>Maes</b> ”)                         |
| <b>1008</b>        | Excerpts from Dafydd Gibbon et al., <i>Handbook of Multimodal and Spoken Dialogue Systems : Resources, Terminology, and Product Evaluation</i> (2000) (“ <b>Gibbon</b> ”)           |
| <b>1009</b>        | Excerpts from <i>Microsoft Computer Dictionary</i> (5th ed. 2002)                                                                                                                   |
| <b>1010</b>        | U.S. Patent No. 6,073,101 to Stephane Herman Maes (filed January 28, 1997, issued June 6, 2000) (“ <b>Maes ’101</b> ”)                                                              |
| <b>1011</b>        | U.S. Patent No. 6,321,196 B1 to Carlos Antonio Franceschi (filed July 2, 1999, issued November 20, 2001) (“ <b>Franceschi</b> ”)                                                    |

**List of Exhibits**

| <b>Exhibit No.</b> | <b>Description of Document</b>                                                                                                                                                     |
|--------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>1012</b>        | U.S. Patent No. 6,829,603 B1 to Joyce Yue Chai, Wlodek Zadrozny et al. (filed February 2, 2000, issued December 7, 2004) (“ <b>Chai</b> ”)                                         |
| <b>1013</b>        | U.S. Patent No. 6,078,914 to Darren M. Redfern (filed December 9, 1996, issued June 20, 2000) (“ <b>Redfern</b> ”)                                                                 |
| <b>1014</b>        | U.S. Patent Application Publication No. 2003/0144846 to Lawrence A. Denenberg and Christopher M. Schmandt (filed January 31, 2002, published July 31, 2003) (“ <b>Denenberg</b> ”) |
| <b>1015</b>        | Excerpts from <i>Merriam-Webster’s Collegiate Dictionary</i> (10th ed. 1994)                                                                                                       |
| <b>1016</b>        | Excerpts from Richard F. E. Sutcliffe et al., <i>Industrial Parsing of Software Manuals</i> (1996) (“ <b>Sutcliffe</b> ”)                                                          |
| <b>1017</b>        | Christopher Schmandt, “The Intelligent Ear: A Graphical Interface to Digital Audio” in Proc. Int’l Conf. Cybernetics and Society, pp.393-397 (1981)                                |
| <b>1018</b>        | Christopher Schmandt et al., “The Intelligent Voice-Interactive Interface,” Proc. Human Factors in Comp. Sys., pp.363-366 (1982)                                                   |
| <b>1019</b>        | Christopher Schmandt et al., “A Robust Parser and Dialog Generator for a Conversational Office System” in Proc. AVIOS Conf. (1986)                                                 |
| <b>1020</b>        | Declaration of Ingrid Hsieh-Yee, Ph.D.                                                                                                                                             |
| <b>1021</b>        | Prosecution history of U.S. Patent No. 8,015,006 B2                                                                                                                                |
| <b>1022</b>        | Proof of Service of Complaint, <i>Dialect, LLC v. Google LLC</i> , Case No. 1:23-cv-00378 (D. Del.)                                                                                |
| <b>1023</b>        | International Preliminary Examination Report for Maes (EX1007) application                                                                                                         |

**List of Exhibits**

**Claim Listing**

|                  | <b>Claim 1</b>                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>1[pre]</b>    | 1. A method for processing natural language speech utterances with context-specific domain agents, comprising:                                                                                                                                                                                                                                                                                                                                 |
| <b>1[a]</b>      | receiving, at a speech unit coupled to a processing device, a natural language speech utterance that contains a request;                                                                                                                                                                                                                                                                                                                       |
| <b>1[b]</b>      | recognizing, at a speech recognition engine coupled to the processing device, one or more words or phrases contained in the utterance using information in one or more dictionary and phrase tables, wherein recognizing the one or more words or phrases contained in the utterance includes:                                                                                                                                                 |
| <b>1[b][i]</b>   | dynamically updating the information in the one or more dictionary and phrase tables based on a dynamic set of prior probabilities or fuzzy possibilities;                                                                                                                                                                                                                                                                                     |
| <b>1[b][ii]</b>  | determining an identity associated with a user that spoke the utterance based on voice characteristics associated with the utterance; and                                                                                                                                                                                                                                                                                                      |
| <b>1[b][iii]</b> | associating the one or more recognized words or phrases and a pronunciation associated with the one or more recognized words or phrases with the determined identity and the request contained in the utterance in response to the one or more recognized words or phrases satisfying a predetermined confidence level;                                                                                                                        |
| <b>1[c]</b>      | parsing, at a parser coupled to the processing device, the one or more recognized words or phrases to determine a meaning associated with the utterance and a context associated with the request contained in the utterance, wherein the one or more recognized words or phrases are further associated with the determined context in response to the one or more recognized words or phrases satisfying the predetermined confidence level; |
| <b>1[d]</b>      | formulating, at the parser, the request contained in the utterance in accordance with a grammar used by a domain agent associated with the determined context;                                                                                                                                                                                                                                                                                 |
| <b>1[e]</b>      | processing the formulated request with the domain agent associated with the determined context to generate a response to the utterance; and                                                                                                                                                                                                                                                                                                    |
| <b>1[f]</b>      | presenting the generated response to the utterance via the speech unit.                                                                                                                                                                                                                                                                                                                                                                        |

### Claim Listing

|        | <b>Claim 2</b>                                                                                                                                                                                                                                                                                   |
|--------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 2[pre] | 2. The method according to claim 1, wherein parsing the one or more recognized words or phrases to determine the context associated with the request includes:                                                                                                                                   |
| 2[a]   | matching the one or more recognized words or phrases to a set of keywords associated with the dynamic set of prior probabilities or fuzzy possibilities;                                                                                                                                         |
| 2[b]   | scoring multiple possible contexts for the one or more recognized words or phrases matched to the set of keywords using a profile associated with the determined user identity, data content associated with the domain agent, or a context stack that includes one or more recent contexts; and |
| 2[c]   | selecting one of the scored multiple possible contexts having a highest score that satisfies the predetermined confidence level to be the determined context.                                                                                                                                    |
|        | <b>Claim 3</b>                                                                                                                                                                                                                                                                                   |
| 3[pre] | 3. The method according to claim 2, wherein parsing the one or more recognized words or phrases to determine the context associated with the request further includes:                                                                                                                           |
| 3[a]   | requesting the user to verify the one or more recognized words or phrases in response to determining that none of the scored multiple possible contexts satisfy the predetermined confidence level; and                                                                                          |
| 3[b]   | determining the context based on information that the user provides in one or more subsequent natural language speech utterances.                                                                                                                                                                |
|        | <b>Claim 4</b>                                                                                                                                                                                                                                                                                   |
| 4[pre] | 4. A method for processing natural language speech utterances with context-specific domain agents, comprising:                                                                                                                                                                                   |
| 4[a]   | receiving, at a speech unit coupled to a processing device, a natural language speech utterance that contains a request;                                                                                                                                                                         |
| 4[b]   | recognizing, at a speech recognition engine coupled to the processing device, one or more words or phrases contained in the utterance using information in one or more dictionary and phrase tables, wherein recognizing the one or more words or phrases contained in the utterance includes:   |

**Claim Listing**

|          |                                                                                                                                                                                                                                                                                                              |
|----------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 4[b][i]  | requesting a user that spoke the utterance to spell the one or more words or phrases in response to the one or more recognized words or phrases failing to satisfy a predetermined confidence level; and                                                                                                     |
| 4[b][ii] | updating the information in the one or more dictionary and phrase tables based on a phonetic alphabet spelling associated with the one or more words or phrases, wherein the user provides the phonetic alphabet spelling in one or more subsequent natural language speech utterances;                      |
| 4[c]     | parsing, at a parser coupled to the processing device, the phonetic alphabet spelling associated with the one or more words or phrases to determine a meaning associated with the utterance and a context associated with the request contained in the utterance;                                            |
| 4[d]     | formulating, at the parser, the request contained in the utterance in accordance with a grammar used by a domain agent associated with the determined context;                                                                                                                                               |
| 4[e]     | processing the formulated request with the domain agent associated with the determined context to generate a response to the utterance; and                                                                                                                                                                  |
| 4[f]     | presenting the generated response to the utterance via the speech unit.                                                                                                                                                                                                                                      |
|          | <b>Claim 5</b>                                                                                                                                                                                                                                                                                               |
| 5[pre]   | 5. A method for processing natural language speech utterances with context-specific domain agents, comprising:                                                                                                                                                                                               |
| 5[a]     | receiving, at a speech unit coupled to a processing device, a natural language speech utterance that contains a request;                                                                                                                                                                                     |
| 5[b]     | recognizing, at a speech recognition engine coupled to the processing device, one or more words or phrases contained in the utterance using information in one or more dictionary and phrase tables;                                                                                                         |
| 5[c]     | parsing, at a parser coupled to the processing device, information relating to the utterance to determine a meaning associated with the utterance and a context associated with the request contained in the utterance, wherein the parsed information includes the one or more recognized words or phrases; |
| 5[d]     | formulating, at the parser, the request contained in the utterance in accordance with a grammar used by a domain agent associated with the                                                                                                                                                                   |

### Claim Listing

|           |                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|-----------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|           | determined context, wherein formulating the request in accordance with the grammar used by the domain agent includes:                                                                                                                                                                                                                                                                                                                                  |
| 5[d][i]   | determining one or more required values and one or more optional values associated with formulating the request in the grammar used by the domain agent;                                                                                                                                                                                                                                                                                               |
| 5[d][ii]  | extracting one or more criteria and one or more parameters from one or more keywords contained in the one or more recognized words or phrases, wherein the parser extracts the one or more criteria and the one or more parameters using procedures sensitive to the determined context;                                                                                                                                                               |
| 5[d][iii] | inferring one or more further criteria and one or more further parameters associated with the request using a dynamic set of prior probabilities or fuzzy possibilities; and                                                                                                                                                                                                                                                                           |
| 5[d][iv]  | transforming the one or more extracted criteria, the one or more extracted parameters, the one or more inferred criteria, and the one or more inferred parameters into one or more tokens having a format compatible with the grammar used by the domain agent, wherein the one or more tokens include all the required values and one or more of the optional values associated with formulating the request in the grammar used by the domain agent; |
| 5[e]      | processing the formulated request with the domain agent associated with the determined context to generate a response to the utterance; and                                                                                                                                                                                                                                                                                                            |
| 5[f]      | presenting the generated response to the utterance via the speech unit.                                                                                                                                                                                                                                                                                                                                                                                |
|           | <b>Claim 6</b>                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| 6[pre]    | 6. A method for processing natural language speech utterances with context-specific domain agents, comprising:                                                                                                                                                                                                                                                                                                                                         |
| 6[a]      | receiving, at a speech unit coupled to a processing device, a natural language speech utterance that contains a request, wherein the request includes at least one command and at least one question;                                                                                                                                                                                                                                                  |
| 6[b]      | parsing, at a parser coupled to the processing device, information relating to the utterance to determine a meaning associated with the utterance and a context associated with the request contained in the utterance;                                                                                                                                                                                                                                |

**Claim Listing**

|      |                                                                                                                                                                                                                                                                                                                                                                           |
|------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 6[c] | formulating, at the parser, the request contained in the utterance in accordance with a grammar used by a domain agent associated with the determined context;                                                                                                                                                                                                            |
| 6[d] | processing the formulated request with the domain agent associated with the determined context to generate a response to the utterance, wherein processing the formulated request with the domain agent includes:                                                                                                                                                         |
| 6[e] | directing the at least one command to one or more local or remote processing devices to execute the at least one command; and                                                                                                                                                                                                                                             |
| 6[f] | querying one or more local or network information sources to resolve information requested in the at least one question, wherein querying the one or more local or network information sources further includes:                                                                                                                                                          |
| 6[g] | submitting multiple asynchronous queries to the one or more local or network information sources, wherein the multiple queries include one or more duplicate queries submitted to distinct information sources;                                                                                                                                                           |
| 6[h] | asynchronously receiving multiple results to the multiple asynchronous queries from the one or more local or network information sources; and                                                                                                                                                                                                                             |
| 6[i] | scoring a relevance associated with the multiple asynchronously received results using a dynamic set of prior probabilities or fuzzy possibilities to determine one or more best responses to the at least one question, wherein scoring the relevance associated with the multiple asynchronously received results to determine the one or more best responses includes: |
| 6[j] | determining one or more tokens required to formulate a response to the at least one question;                                                                                                                                                                                                                                                                             |
| 6[k] | applying one or more scraping criteria to the multiple asynchronously received results to extract one or more values for the asynchronously received results;                                                                                                                                                                                                             |
| 6[l] | evaluating the one or more extracted values using the dynamic set of prior probabilities or fuzzy possibilities to resolve ambiguous, incomplete, or conflicting information associated with the one or more extracted values; and                                                                                                                                        |
| 6[m] | selecting one or more best values for the one or more required tokens from the one or more evaluated values, wherein the one or more best                                                                                                                                                                                                                                 |

**Claim Listing**

|        |                                                                                                                                                                                                                                                                                                                   |
|--------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|        | responses include the one or more best values for the one or more required tokens; and                                                                                                                                                                                                                            |
| 6[n]   | presenting the generated response to the utterance via the speech unit, wherein the response presented via the speech unit includes results associated with the one or more local or remote processing device executing the at least one command and the one or more best responses to the at least one question. |
|        | <b>Claim 9</b>                                                                                                                                                                                                                                                                                                    |
| 9[pre] | 9. The method according to claim 6, wherein scoring the relevance associated with the multiple asynchronously received results to determine the one or more best responses further includes:                                                                                                                      |
| 9[a]   | determining that the dynamic set of prior probabilities or fuzzy possibilities failed to resolve the ambiguous, incomplete, or conflicting information associated with the one or more extracted values; and                                                                                                      |
| 9[b]   | requesting a user that spoke the utterance to provide additional information relating to the request, wherein the domain agent uses the additional information relating to the request to resolve the ambiguous, incomplete, or conflicting information associated with the one or more extracted values.         |

**I. MANDATORY NOTICES UNDER §42.8(A)(1)**

**A. Real Party-In-Interest under §42.8(b)(1)**

Google LLC (“Petitioner”) is the real party-in-interest to this IPR petition.<sup>1</sup>

**B. Related Matters under §42.8(b)(2)**

The ’006 patent is the subject of two currently-pending litigations: *Dialect, LLC v. Google LLC*, Case No. 1:23-cv-00378 (D. Del.) (filed April 3, 2023), in which Petitioner is named as a defendant, and *Dialect, LLC v. Amazon.com, Inc.*, Case No. 1:23-cv-00581-TSE-LRV (E.D. Va.) (filed May 1, 2023). Petitioner was first served with a complaint alleging infringement of the ’006 patent on April 6, 2023. (EX1022, p.002.) On March 28, 2024, Petitioner’s motion to transfer to the Northern District of California was granted. *See* Dkt. 32, Case No. 1:23-cv-00378 (D. Del.).

Petitioner is filing IPR petitions against eight patents asserted by Patent Owner in the pending litigation involving Petitioner:

| U.S. Patent No. | IPR Case      |
|-----------------|---------------|
| 7,398,209 B2    | IPR2024-00746 |
| 7,502,738 B2    | IPR2024-00747 |
| 8,015,006 B2    | IPR2024-00748 |

---

<sup>1</sup> Google LLC is a subsidiary of XXVI Holdings Inc., which is a subsidiary of Alphabet Inc. XXVI Holdings Inc. and Alphabet Inc. are not real parties in interest to this proceeding.

Petition for *Inter Partes* Review of  
U.S. Patent No. 8,015,006 B2

| U.S. Patent No. | IPR Case      |
|-----------------|---------------|
| 7,693,720 B2    | IPR2024-00749 |
| 9,031,845 B2    | IPR2024-00750 |
| 8,447,607 B2    | IPR2024-00751 |
| 8,849,652 B2    | IPR2024-00752 |
| 7,640,160 B2    | IPR2024-00753 |

**C. Lead and Back-Up Counsel under §42.8(b)(3)**

Petitioner provides the following designation of counsel.

| LEAD COUNSEL                                                                                                                                                                                                | BACK-UP COUNSEL                                                                                                                                                                                                                                       |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Heidi L. Keefe (Reg. No. 40,673)<br>hkeefe@cooley.com<br><br>COOLEY LLP<br>ATTN: Patent Group<br>1299 Pennsylvania Ave. NW, Suite 700<br>Washington, DC 20004<br>Tel: (650) 843-5001<br>Fax: (650) 849-7400 | Andrew C. Mace (Reg. No. 63,342)<br>amace@cooley.com<br><br>COOLEY LLP<br>ATTN: Patent Group<br>1299 Pennsylvania Ave. NW, Suite 700<br>Washington D.C. 20004<br>Tel: (650) 843-5287<br>Fax: (650) 849-7400                                           |
|                                                                                                                                                                                                             | Mark R. Weinstein (Admission <i>pro hac vice</i> to be requested)<br>mweinstein@cooley.com<br><br>COOLEY LLP<br>ATTN: Patent Group<br>1299 Pennsylvania Avenue NW<br>Suite 700<br>Washington D.C. 20004<br>Tel: (650) 843-5007<br>Fax: (650) 849-7400 |
|                                                                                                                                                                                                             | Reuben Chen (Admission <i>pro hac vice</i> to be requested)<br>rchen@cooley.com<br><br>COOLEY LLP                                                                                                                                                     |

| LEAD COUNSEL | BACK-UP COUNSEL                                                                                                                       |
|--------------|---------------------------------------------------------------------------------------------------------------------------------------|
|              | ATTN: Patent Group<br>1299 Pennsylvania Avenue NW<br>Suite 700<br>Washington D.C. 20004<br>Tel: (650) 843-5007<br>Fax: (650) 849-7400 |

**D. Service Information**

This Petition is being served by Federal Express to the correspondence address of record for the '006 patent, 174899 - The Law Offices of David A. Gerasimow, P.C., 211 W. Wacker Dr., Ste. 1717, Chicago, IL 60606. Petitioner consents to electronic service at the addresses provided above for lead and back-up counsel.

**II. FEE PAYMENT**

Petitioner requests review of 7 claims, with a \$41,500 payment.

**III. REQUIREMENTS UNDER §§ 42.104 AND 42.108 AND CONSIDERATIONS UNDER §§ 314(A) AND 325(D)**

**A. Grounds for Standing**

Petitioner certifies that the '006 patent is available for IPR and that Petitioner is not barred or otherwise estopped.

**B. Identification of Challenge and Statement of Precise Relief Requested**

Petitioner requests IPR institution of claims 1-6 and 9 based on:

| Ground | Claims | Basis for Challenge under §103                                          |
|--------|--------|-------------------------------------------------------------------------|
| 1      | 1-3    | Coffman (EX1003), Kanevsky (EX1004), Kneser (EX1005), Zadrozny (EX1006) |
| 2      | 4      | <b>Ground 1</b> Prior Art + Maes '101 (EX1010) and Franceschi (EX1011)  |
| 3      | 5      | <b>Ground 1</b> Prior Art + Chai (EX1012)                               |
| 4      | 6, 9   | <b>Ground 1</b> Prior Art + Redfern (EX1013)                            |

### C. Considerations Under §§ 314(a) and 325(d)

**§314(a):** Petitioner is unaware of another post-grant challenge against the '006 patent. The pending litigation involving Petitioner is still in early stages, with no discovery or claim construction having occurred and no trial date set, and as noted, was recently transferred to the Northern District of California. If IPR is instituted, Petitioner intends to bring a motion to stay the pending litigation.

**§325(d):** During '006 patent prosecution, an applicant-submitted IDS identified Coffman's U.S. counterpart (i.e., 7,137,126) and Chai. (EX1021, pp.00123-00124.) Neither was substantively discussed. Why the Examiner did not appreciate their significance is unclear, but they were listed among approximately 300 other references. (EX1021, pp.00117-00128.) The Examiner also did not have the additional references that this Petition relies on, which provide further context and detail for understanding Coffman and how its disclosures provide compelling grounds of obviousness.

#### **IV. OVERVIEW OF THE CHALLENGED PATENT**

##### **A. Person of Ordinary Skill in the Art (POSITA)**

A POSITA as of June 3, 2002<sup>2</sup> would have possessed a bachelor's degree in electrical engineering or computer science, and two years of work experience in the field of speech- and natural language-interfaces for interacting with and directing computer systems and software, including speech recognition and processing (and more experience could substitute for less formal education, and vice versa). (EX1002 (“Schmandt”), ¶¶11-15.)

##### **B. Specification Overview**

The '006 patent states that it “relates to the retrieval of online information or processing of commands through a speech interface,” as illustrated:

---

<sup>2</sup> Petitioner does not concede that the claims are entitled to the June 3, 2002 date, but that issue is irrelevant here because the prior art predates that date.

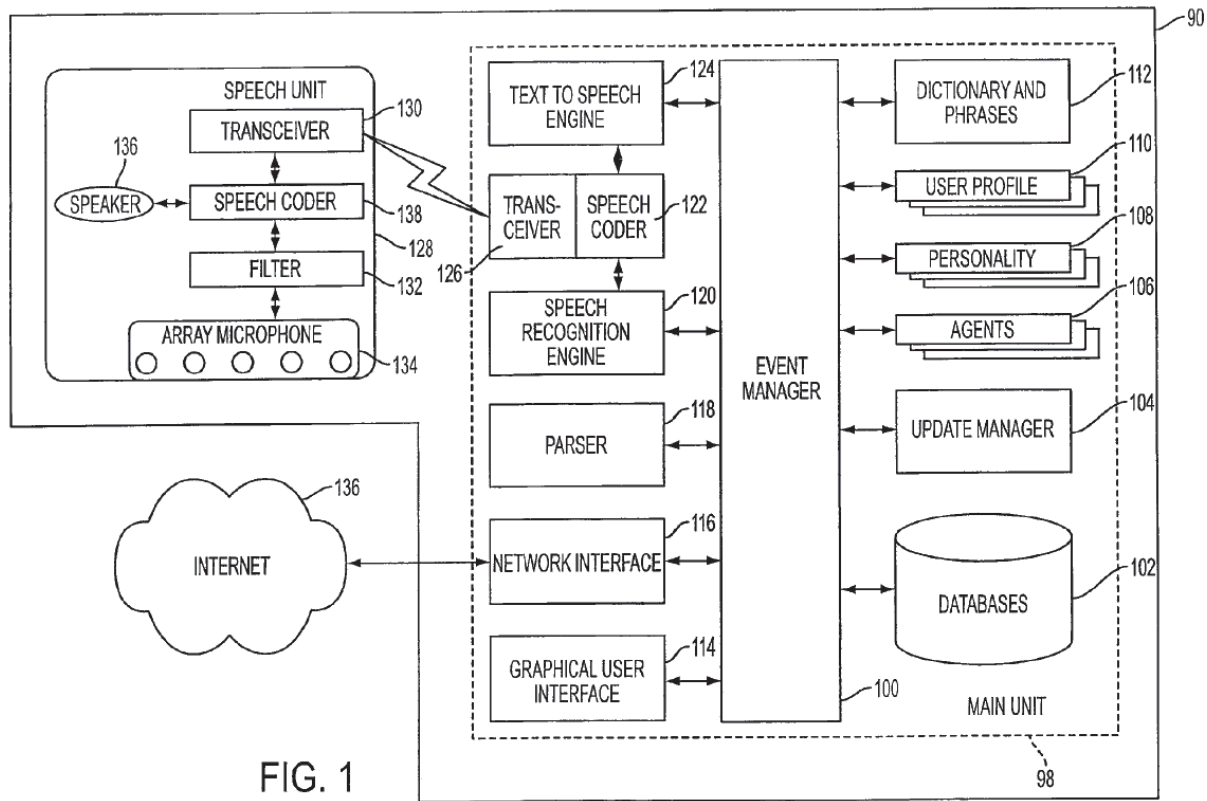


FIG. 1

(’006, Fig. 1, 1:20-22.) As shown, system **90** includes speech unit **128**, speech recognition engine **120**, dictionary and phrases **112**, parser **118**, and agents **106**. Speech unit **128** includes a microphone to receive a spoken utterance. (’006, 10:30-32.) The speech is passed to speech recognition engine **120** for processing “using data in the dictionary and phrases module **112** and received from the agents **106**.” (’006, 10:56-60.) Then, “[t]he recognized words and phrases are processed by the parser **118**, which transforms them into complete commands and questions using data supplied by an agent[s] **106**.” (’006, 10:61-63.) “The agent[s] **106** then process the commands or questions.” (’006, 10:63-64.) To select the correct agent to invoke to process the command or question, “parser **118** determines a context for an

utterance by applying prior probabilities or fuzzy possibilities to keyword matching, user profile **110**, and dialog history.” (’006, 17:49-51, 18:12-14.)

## **V. CLAIM CONSTRUCTION**

Express claim construction is unnecessary because the prior art renders the challenged claims obvious under any reasonable construction. Petitioner reserves its right to respond to claim construction arguments by Patent Owner.

## **VI. THE CHALLENGED CLAIMS ARE UNPATENTABLE**

### **A. Overview of Grounds**

The challenged claims attempt to lay claim to well-known techniques for speech-enabled interfaces for interacting with and directing computer systems. The challenged claims employ standard techniques that were known long before the ’006 patent, such as using speech recognition and natural language processing to select an application for carrying out a query and/or command included in a spoken utterance.

All grounds rely primarily on **Coffman** (**EX1003**), which despite being filed nearly three years before the ’006 patent, discloses a remarkably similar conversational computing system. All grounds combine Coffman with **Kanevsky** (**EX1004**), **Kneser** (**EX1005**) and **Zadrozny** (**EX1006**), which disclose known techniques related to speech recognition and natural language processing. Grounds 2-4 combine these references with **Maes ’101** (**EX1010**), **Franceschi** (**EX1011**),

**Chai (EX1012)** and/or **Redfern (EX1013)** for features recited in claims 2-6 and 9.

As detailed below, these references share links to each other and thus comprise a collective body of work describing aspects of the same systems and methods—for example, Coffman expressly incorporates Kanevsky by reference (Coffman, 37:10-13), and Kanevsky in turn discusses Zadrozny (Kanevsky, 14:22-25); Coffman, Kanevsky, and Maes '101 share a common named co-inventor (Stephane Maes), as do Zadrozny and Chai (Wlodek Zadrozny); and the named inventors and authors of Coffman, Kanevsky, Zadrozny, Maes '101, Franceschi and Chai are identified as associated with IBM. *See, e.g., Paice, LLC v. Ford Motor Co.*, 881 F.3d 894, 906-907 (Fed. Cir. 2018) (incorporated material “effectively part of the host document as if it were explicitly contained therein”); *see also Boston Scientific Scimed, Inc. v. Cordis Corp.*, 554 F.3d 982, 991 (Fed. Cir. 2009) (“[c]ombining two embodiments disclosed adjacent to each other in a prior art patent does not require a leap of inventiveness”).

The references are prior art under at least pre-AIA §§102(b) (Coffman, Kanevsky, Kneser, Zadrozny, Maes '101, Redfern), 102(a) (Franceschi) and 102(e) (Chai). (*See* EX1020, ¶¶25-48 (Zadrozny).)

**B. Ground 1: Claims 1-3 Are Obvious Over Coffman in View of Kanevsky, Kneser and Zadrozny.**

**1. Claim 1: “A method...”**

As shown below, the claimed method is obvious over Ground 1 prior art.

(a) “receiving...” (1[a])

The '006 patent describes a “**speech unit**” as comprising a microphone and speaker, with other optional elements, and connected to a “main unit” processing device. ('006, 10:28-39, 10:51-53, Fig. 1; Schmandt, ¶83.)

Coffman’s system includes “**a speech unit coupled to a processing device**”—specifically a microphone and speaker for receiving a user’s speech input and providing audio output (e.g., via text-to-speech) (Schmandt, ¶84; Coffman, 53:8-10 (“Voice input from a microphone”), 37:14-22, 35:2-3 (“listen to the output of a task”), 39:12 (“audio only output”), 39:30 (“voice output”), Fig. 6 (telephone 600, “speech synthesis,” “speech out”).)<sup>3</sup> These components are coupled to components for processing input from and output to them (“**processing device**”), such as conventional processing components (e.g., CPU) of a desktop computer or PDA for running a conversational virtual machine (“CVM”) and related software. (Schmandt, ¶84; Coffman, 10:26-28, 12:5-7, 13:1-3, 28:29-31, 20:26-29 (describing prioritizing “CPU cycles”), 45:29-32 (“CPU load” and “CPU cycles”), Figs. 1 & 6.)

Coffman further discloses receiving a person’s spoken input using, e.g., a telephone or microphone (“**receiving, at a speech unit coupled to a processing device, a...speech utterance**”). (Coffman, 53:8-10, 37:14-22, 41:2-5, 53:22-26.)

Coffman illustrates:

---

<sup>3</sup> All underlining added to quoted material unless noted.

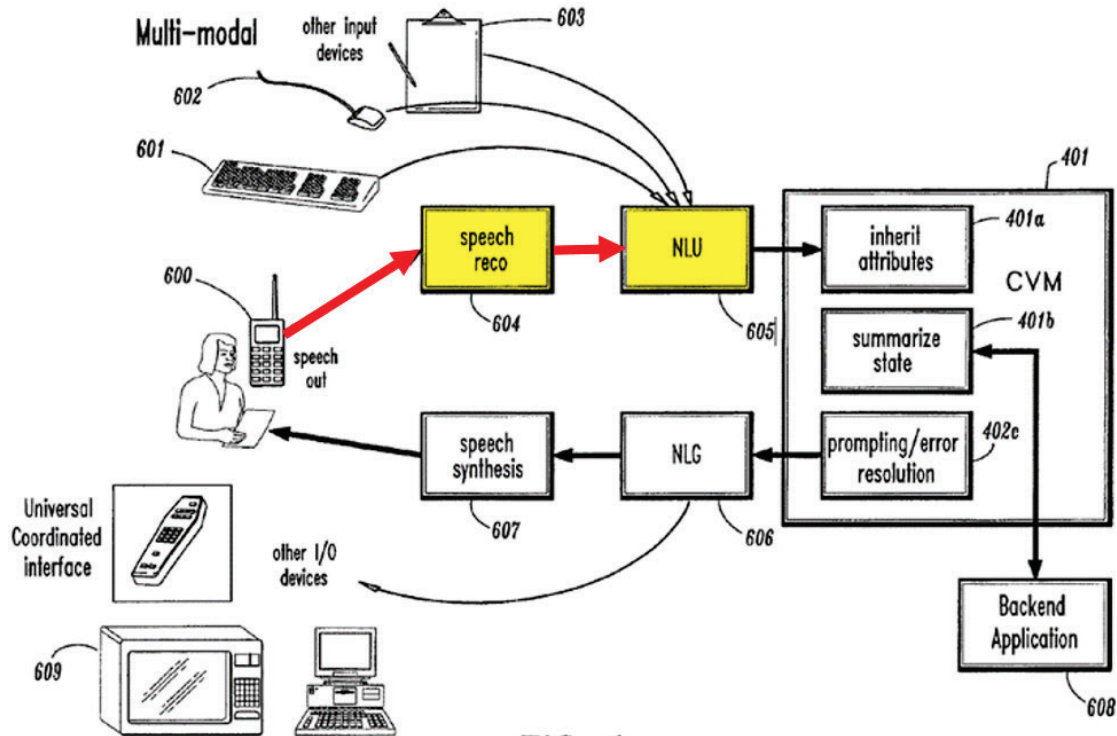
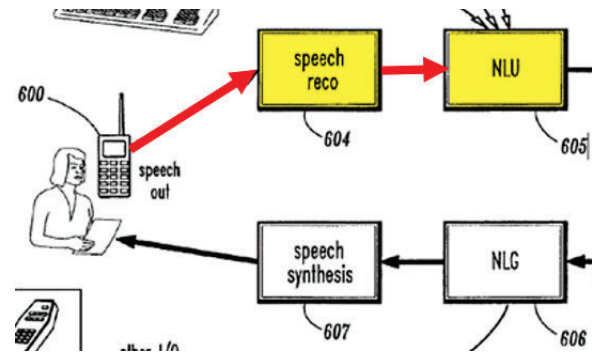


FIG. 6

(Coffman, Fig. 6 (annotations added); *see also id.*, 37:14-22 (“Spoken input from a speech client (e.g., telephone 600)...”).)

Coffman’s spoken input is a “**natural language speech utterance.**” In the art, “**natural language**” generally refers to the forms of language used by a person for communication (e.g., English). (Schmandt, ¶86 (citing EX1008, p.0035 (“natural language”)).) Coffman discloses several examples of natural language speech utterances. (Coffman, e.g., 6:32-7:3 (“the command ‘I would like to open this’”), 48:1-3 (“e-mail from my mother”), 56:25-7:2 (“grocery list,” “it should be synchronized with my PDA,” and “Do it!”).)

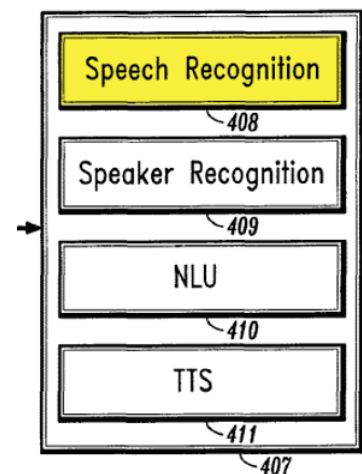
Receipt of spoken “**natural language**” input is confirmed by Coffman’s explanation that spoken input, once recognized, is subjected to natural language parsing and understanding (“NLU”) processes. (Coffman, Fig. 6 (annotated excerpt at right), 53:22-26 (“parsing and NLU”), 53:8-10 (“symbolic language representation provided by NLU”), 37:14-22 (“NLU processing 605”), 6:6-8 (“natural dialog with NLU”), 37:29-32, 5:1-4; Schmandt, ¶87.)



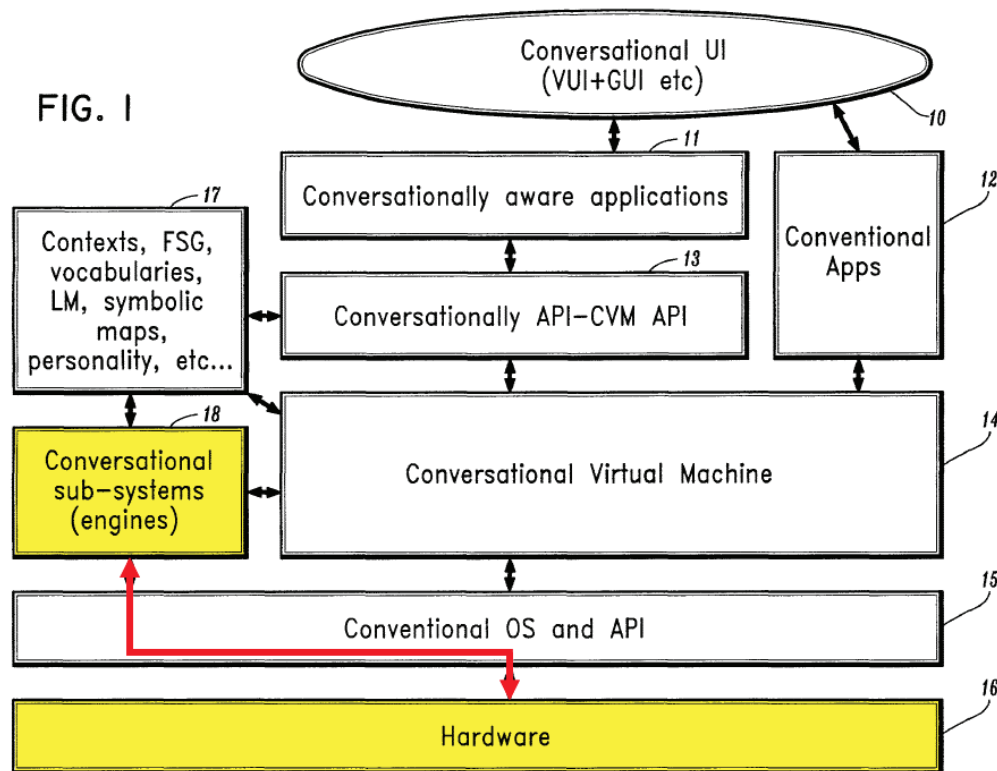
Coffman further discloses and renders obvious that the speech utterance “**contains a request,**” explaining that the system “convert[s] voice requests into queries.” (Coffman, 13:8-13.) Coffman’s example “I would like to open this,” is a spoken utterance that “**contains a request.**” (Schmandt, ¶88.)

(b) “recognizing...” (1[b])

Coffman’s system includes various “conversational subsystems,” including a “**speech recognition engine**” (illustrated at right (Fig. 4 (excerpt))). (Coffman, 10:28-30 (“...conversational subsystems (e.g. speech recognition engine, text-to speech etc.)...”), 29:1-4 (“...a plurality of conversational resource subsystems (engines) 407 including, for example, a speech recognition system 408”), Fig. 4.) A POSITA



would have understood that Coffman’s speech recognition engine, as a software component running on a conventional device such as a desktop computer or PDA, would be “**coupled to the processing device.**” (Schmandt, ¶89; Coffman, 13:1-3, 20:26-29, 45:29-32, Figs. 1 & 6.) For example, Coffman Figure 1 depicts conversational sub-systems (engines) 18—which includes a “speech recognition engine”—coupled to hardware 16:



(Coffman, Fig. 1 (annotations added); *see also id.*, 13:1-3, 29:1-4.)

Coffman’s speech recognition engine “**recogniz[es] one or more words or phrases contained in the utterance,**” explaining that “[s]poken input from a speech client (e.g., telephone 600) is subject to a speech recognition process 604.”

(Coffman, 37:20-22; *see also id.*, Fig. 6 (depicting input to and output from “speech reco” 604), 38:9-11.)

A POSITA would have also understood and found it obvious that the “**recognizing**” in Coffman’s system would be performed “**using information in one or more dictionary and phrase tables**”—e.g., n-gram language models used for speech recognition. (Schmandt, ¶91.) Coffman explains that each conversationally-aware application registers with a conversational virtual machine (“CVM”), and provides its language model(s) for controlling the application. (Coffman, 57:22-58:11, 7:12-14, 13:8-13.) In a speech recognition context, a “language model” generally refers to a formal representation of the structure of a natural language. (Schmandt, ¶92 (citing EX1008, p.0031 (“language model” definition)).) “In most speech recognisers language models consist of an n-gram model, i.e. probabilities of the occurrence of words, words pairs, sequences of three words, etc.” (*Id.* (quoting EX1008, p.0031; citing EX1008, pp.0035 (“n-gram language model”), 0050 (“stochastic language model”))).) A “trigram” language model, for example, therefore refers to an n-gram language model that uses the probabilities of the occurrence of sequences of three words (i.e., the “n” in “n-gram” is three). (*Id.* (citing EX1008, p.0053 (“trigram”))).)

The use of language models for speech recognition was well-known. (Schmandt, ¶93.) *See, e.g., Bayer Healthcare LLC v. Baxalta Inc.*, 989 F. 3d 964,

982 (Fed. Cir. 2021) (“[a] patent need not teach, and preferably omits, what is well known in the art”). And Coffman expressly incorporates **Kanevsky** (EX1004) (Coffman, 37:10-13), which further explains that a speech recognition engine may be implemented using a “trigram language model.” (Kanevsky, 13:50-56; *see also id.*, 13:48-61 (describing such a speech recognition engine as an “existing component”).)

Under the combination of Coffman and Kanevsky, the speech recognition process to identify trigram(s) corresponding to the utterance discloses **“recognizing...one or more words or phrases contained in the utterance”** because the trigrams comprise words and sequences of words (i.e., phrases). (Schmandt, ¶94.) Moreover, recognizing trigrams from the user’s spoken input under the combination **“us[es] information in one or more dictionary and phrase tables”** because a POSITA would have understood and found it obvious to organize the n-gram language model(s) as a **“table.”** (Schmandt, ¶94 (citing EX1016, p.006 (with n-gram language model, “syntactic knowledge is captured in a large table of word bigrams (pairs), trigrams (triples),...N-grams”); Denenberg, ¶0040 (“...N-gram language model [] works from tables of probabilities of sequences of words”))).) Thus, a trigram language model comprises **“one or more dictionary and phrase tables”** because it is organized as a table and includes a list of words (**“dictionary”**) and sequences of words (**“phrase(s)”**) to refer to during speech

recognition. (*Id.* (citing EX1015, p.005 (“dictionary” definition encompassing “a list (as of items of data or words) stored in a computer for reference (as for information retrieval or word processing)”).))

**Rationale and Motivation to Combine (Coffman and Kanevsky).** It would have been obvious to combine Coffman and Kanevsky, predictably resulting in Coffman’s conversational system in which language models provided by the conversationally-aware applications and used by the system for speech recognition were n-gram language models (e.g., trigram language models). (Schmandt, ¶95.) Combining their teachings in this manner would have resulted in “**recognizing, at a speech recognition engine coupled to the processing device, one or more words or phrases contained in the utterance using information in one or more dictionary and phrase tables.**” (*Id.*)

Coffman and Kanevsky are analogous references in the field of the ’006 patent, and would have been reasonably pertinent to problems facing the patent inventors, including how to implement a system for a person to interact with a computer system using speech and techniques for performing speech recognition.. (Schmandt, ¶96 (citing ’006, 1:20-22; Coffman, 1:7-8, Figs. 1, 4 & 6; Kanevsky, 1:15-21).)

A POSITA would have been motivated to combine. Coffman expressly incorporates Kanevsky (Coffman, 37:10-13) and is therefore part of Coffman’s

disclosures for purposes of applying it as prior art. *See, e.g., Paice*, 881 F.3d 894 at 906-907 (“...effectively part of the host document as if it were explicitly contained therein”); *see also Boston Scientific Scimed*, 554 F.3d at 991 (“[c]ombining two embodiments disclosed adjacent to each other in a prior art patent does not require a leap of inventiveness”). Moreover, both references share a named co-inventor, Stephane Maes, and identify IBM as the assignee/applicant. (Coffman, Kanevsky face pages.) Therefore, a POSITA would have consulted both of these references to understand the features and functionality of Coffman’s conversational system and/or to actually implement or adapt the system and features it provides. (Schmandt, ¶¶97.)

Moreover, Coffman already discloses that a conversationally-aware application provides its “language models” to the conversational system in connection with processing a user’s input. A POSITA would have understood that a language model in this context likely refers to an n-gram language model, and Kanevsky provides express confirmation, explaining that a speech recognition engine can be implemented using a trigram language model. (Schmandt, ¶¶98 (citing EX1008, pp.0031 (“language model”), 0035 (“n-gram language model”); Kanevsky, 13:50-56).) Moreover, using an n-gram language model as Kanevsky teaches would have involved a known technique to implement or improve Coffman’s speech recognition functionality in the same way. (*Id.*)

A POSITA would have had a reasonable expectation of success, as the combination would have involved applying well-known technologies and techniques without undue experimentation. (Schmandt, ¶99 (citing '006, 1:45-48; Kanevsky, 13:48-61).)

(i) “dynamically updating...” (1[b][i])

1[b][i] describes nothing more than conventional use of an “adaptive” n-gram language model for speech recognition. (Schmandt, ¶101.)

An n-gram language model (e.g., trigram language model) includes a “**set of prior probabilities**”—the set of probabilities associated with the trigrams in the trigram language model used by the speech recognition engine and maintained by the conversational system as a result of application registration. (Schmandt, ¶103.) A POSITA would have understood and found it obvious that the probabilities associated with the language model trigrams are “**prior probabilities**” because they each correspond to the probability of observing some phenomenon estimated from previously collected training data, independent of future observations. (*Id.* (citing EX1008, p.0043 (“prior probability”), EX1016, p.006 (explaining that a language model is “extracted or learnt from a corpus, a representative set of texts in the domain of the speech recognizer.”))).)

A “**dynamic set of prior probabilities**” is disclosed and rendered obvious through the conventional use of an “adaptive” n-gram language model for speech

recognition, where the probabilities associated with the n-grams are dynamically updated with use. (Schmandt, ¶104 (citing EX1008, p.0024 (“adaptive language model”))). Adaptive n-gram language models and their advantages were well known, and their use would have been obvious over Kneser (EX1005). (Schmandt, ¶104; Kneser, 1:22-32; *see generally* Kneser, 1:13-37.)

**Rationale and Motivation to Combine (Coffman and Kanevsky with Kneser)**. It would have been obvious to further combine Coffman and Kanevsky with Kneser’s teachings regarding an adaptive language model, predictably resulting in language models in Coffman’s conversational system having a “**dynamic**” set of prior probabilities. (Schmandt, ¶105.) Therefore, under the combination, because the prior probabilities are included in the trigram language model(s) (“**the dictionary and phrase tables**”), 1[b][i] simply corresponds to dynamically updating the trigram language model(s) in accordance with model adaptation as taught by Kneser. (*Id.*)

Kneser is an analogous reference in the field of the ’006 patent, and would have been reasonably pertinent to problems facing the ’006 inventors, including techniques for performing speech recognition and how to implement a speech recognition system that uses a “dynamic” set of prior probabilities. (Schmandt, ¶106.)

Kneser provides express motivation for incorporating “adaptive” language model techniques into Coffman’s conversational system—improving speech recognition reliability and performance. (Schmandt, ¶107.) Coffman itself confirms this motivation in the context of collecting user-specific information to perform “more robust recognition.” (*Id.* (citing Coffman, 37:5-10).) Adaptive language models as taught by Kneser provide a way in which Coffman’s system could advantageously improve speech recognition reliability and performance. Adaptive language models were well known, and a POSITA would have appreciated that using adaptive language models in Coffman’s conversational system for speech recognition would have improved a similar system in the same way. (*Id.*)

A POSITA would have had a reasonable expectation of success, as the combination would have involved applying well-known technologies and techniques without undue experimentation. (Schmandt, ¶108.)

**(ii) “determining...” (1[b][ii])**

Coffman discloses “**determining an identity of the user**” because its system includes functionality for “speaker recognition” and “speaker identification,” such as “text-independent speaker identification” using a user’s command or request:

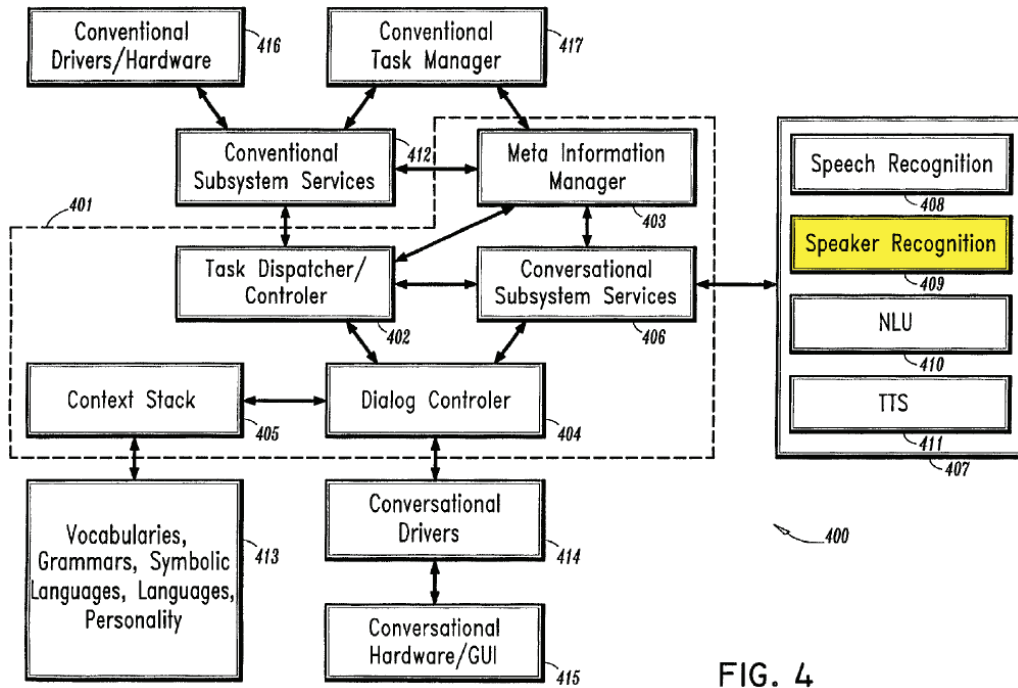


FIG. 4

(Coffman, Fig. 4 (highlighting added); *see also id.*, 5:1-4 (“speaker recognition”), 29:1-4, 15:13-16:24 (listing speaker-identification-related “conversational foundation classes” including “Acoustic\_Identify\_speaker,” “Acoustic\_Verify\_speaker” and “Speech\_biometrics\_identify.”).) Coffman further discloses that “[i]n a speech-based system, each command can be used to perform text-independent speaker identification” (Coffman, 47:25-27; *see also id.*, 8:16-18), which involves using speech signal features to determine the speaker’s identity and does not require the utterance to comprise specific linguistic content. (Schmandt, ¶110 (citing EX1008, pp.0047 (“speaker recognition”), 0051 (“text-independent speaker recognition system”))).)

Kanevsky, which Coffman specifically cites for speaker identification (Coffman, 37:5-13), further describes text-independent speaker recognition that involves comparing “**voice characteristics**” of the user’s utterance with a model to verify the user’s identity. (Kanevsky, 6:66-7:8, 6:12-15 (“The identity claim may be made from...voice characteristics...”), 2:6-8 (describing prior art systems that “employ voice-based verification” using “acoustic characteristics of the utterances spoken by the user”).) The rationale and motivation to combine Coffman and Kanevsky in the analysis of 1[a] applies here also. 1[b][ii] therefore would have been obvious. (Schmandt, ¶111.)

**(iii) “associating...” (1[b][iii])**

Coffman’s system includes a “context stack” where the “context,” including the state and history of each active process (such as an application process) is tracked. (Coffman, 29:18-30:3, 30:4-10.) The context stack is searched to determine a context that matches the user’s request based on NLU processing. (Coffman, 42:1-5, 41:2-5, 30:8-10.)

In Coffman, each context can be associated with a user identity and with the user’s request. With respect to “**the identity of the user**,” Coffman explains that “user identity can be associated in the context of each discourse,” so that “command ambiguity can be immediately and transparently resolved (e-mail from my mother is correctly understood, independently of the user).” (Coffman, 48:1-3; *see also id.*,

46:26-31 (“the stack of contexts should contain the identity of the user as the most recently authenticated identity”), 48:11-14.) With respect to “**the request contained in the utterance,**” the recognized words and phrases are used to match a context and its associated formulated request (e.g., in symbolic language or application call/query). (Coffman, 42:1-7, 37:2-4.) Therefore, as a result of identifying a context, “**the one or more recognized words or phrases**” are “**associated**” with the user’s identity and the request associated with the context (“**with the determined identity and the request contained in the utterance**”). (Schmandt, ¶114.)

1[b][iii] further recites “**associating...a pronunciation associated with the one or more recognized words or phrases with the one or more recognized words or phrases with the determined identity and the request contained in the utterance.**” This would have been obvious over the combination in at least two separate ways. **First**, Coffman explains that application registration involves providing baseforms (pronunciation rules) for the active vocabulary to the conversational system. (Coffman, 57:22-27, 60:30-61:2, 24:30-25:3.) A POSITA would have therefore understood that successful recognition of words and phrases would have involved successfully matching the user’s utterance to the pronunciation rules. (Schmandt, ¶115.) Therefore, the subsequent matching of the recognized words and phrases to a context would also result in associating the pronunciations

specified by the baseforms of those words and phrases with the user's identity and the request associated with the context. (*Id.*)

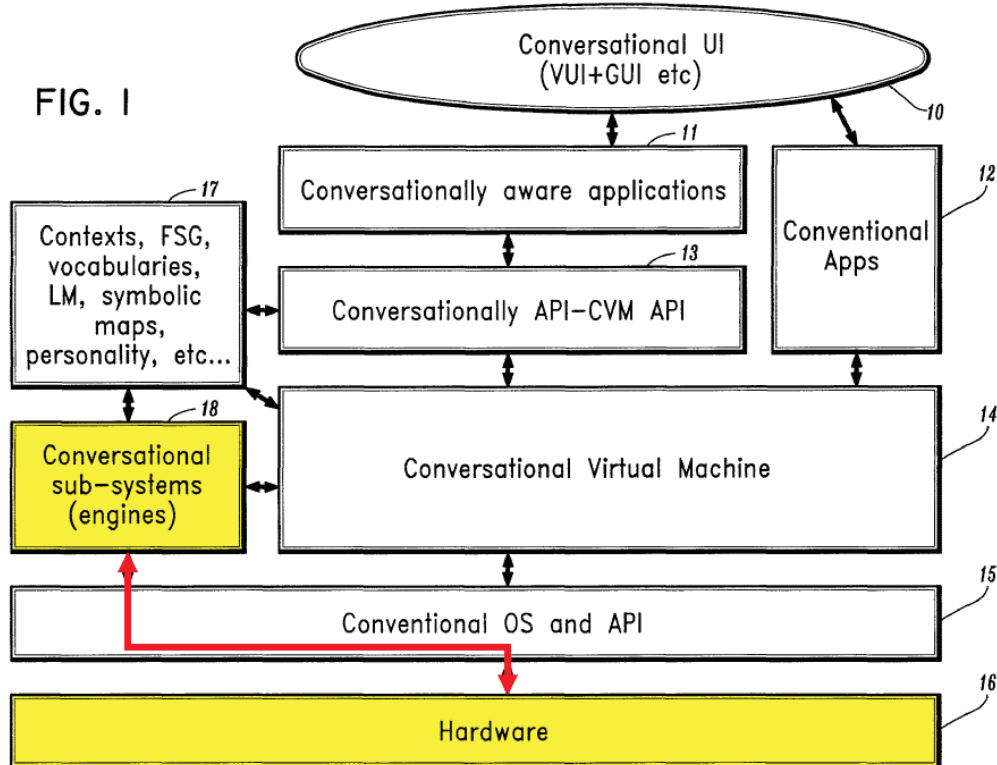
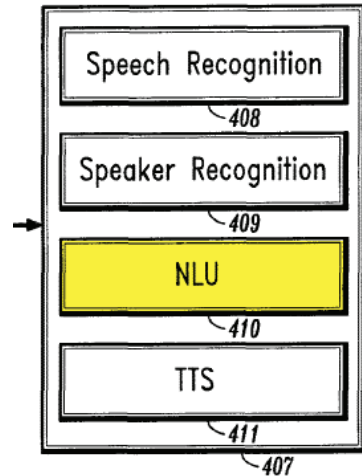
**Second**, as explained for 1[b][ii], identifying the user based on voice characteristics of the user's utterance is disclosed by and obvious over Coffman and Kanevsky. A POSITA would have understood that a user's pronunciation of words and phrases would be reflected in the voice characteristics of their utterance. (Schmandt, ¶116 (citing Kanevsky, 9:39-42 ("prosody" & "accent"), 10:34-38, 11:53-58; EX1008, pp.0056 (defining "voice characteristic" as including "dialect"), 0044 ("prosody"); EX1015, p.004 (defining "dialect" as including "pronunciation"))). Therefore, a POSITA would have understood and found it obvious that, by performing speaker identification in accordance with 1[b][ii], the user's pronunciation of the recognized words and phrases would have been **"associat[ed]...with the determined identity and the request contained in the utterance."** (*Id.*)

It also would have been obvious over Coffman and Kanevsky for the **"associating"** to occur **"in response to the one or more recognized words or phrases satisfying a predetermined confidence level."** As Petitioner's expert explains, the concept of a confidence level has been well known in the art since at least the 1970s. (Schmandt, ¶117 (citing EX1017, pp.002-003).) Coffman itself mentions "confidence and ambiguity results," applications providing a "confidence

level” for its interpretation of an input, and “provid[ing] feedback to the user for incomplete or ambiguous queries/commands.” (Coffman, 27:16, 31:26-28, 40:16-17). It would have been obvious in view of Kanevsky to perform the claimed “**associating**” under the combination only when speaker identification based on the recognized words and phrases satisfied a predetermined confidence level. (Schmandt, ¶117.) Doing so would help ensure that the system did not incorrectly associate an input with a particular user, context and query and would further the security objectives described in both Coffman and Kanevsky. (*Id.* (citing Coffman, 37:5-14 (describing conversational biometrics for query authorization), 8:10-15 (describing preventing unauthorized access), 46:17-32 (describing user authentication and authorization)).) For example, Kanevsky describes calculating a score from text-independent speaker recognition, comparing the score to “a predetermined threshold value,” and allowing access only when the score is “above or within an acceptable predetermined range of the threshold value.” (Kanevsky, 6:66-7:24.) A POSITA would have understood that a score would be compared to a “**predetermined confidence level**” because the score would reflect the system’s confidence whether the context was acceptable. (Schmandt, ¶117.)

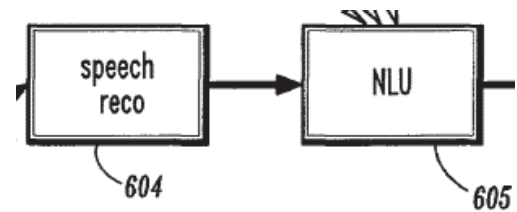
(c) “parsing...” (1[c])

As noted, Coffman explains that its system includes various “conversational subsystems,” including “natural language understanding and natural language parsing system [NLU] 410” (illustrated at right (Fig. 4 (excerpt))). (Coffman, 29:1-4, Fig. 4.) A POSITA would have understood that Coffman’s natural language parser, as a software component running on a conventional device such as a desktop computer or PDA, would be “coupled to the processing device,” as Coffman Figure 1 depicts:



(Coffman, Fig. 1 (annotations added); Schmandt, ¶118 (citing Coffman, 13:1-3,

A POSITA would have understood and found it obvious that Coffman's NLU 410 would **"pars[e]...the one or more recognized words or phrases"** as part of performing natural language understanding. (Schmandt, ¶119.) Coffman repeatedly describes "NL parsing" and "NLU" and depicts the output of speech recognition (**"recognized words or phrases"**) as input to NLU (see Fig. 6 excerpt at right). (Coffman, 29:1-6, 19:16-19, Fig. 4 (NLU 410), Fig. 6 ("speech reco" 604, NLU 605).) Moreover, the incorporated Kanevsky reference explains that NLU can be implemented using "a combination of classical parsing capabilities with key word spotting and speech recognition." (Kanevsky, 14:1-7.)



Coffman further teaches that parsing the recognized words and phrases as part of natural language understanding is **"to determine a meaning associated with the utterance and a context associated with the request contained in the utterance."** This is because Coffman uses NLU in determining a "context" for the user's request to properly carry it out. As discussed for 1[b][iii], Coffman's conversational system includes a "context stack" where the state and history of each active process, such as an application process, is tracked in its own "context." (Coffman, 29:18-30:3; *see also id.*, 30:4-10.) For 1[c], Coffman further explains that the context stack is searched to determine a context that matches the user's request based on NLU

processing. (Coffman, 42:1-5, 41:2-5, 30:8-10.) According to an example from Coffman, using NLU, Coffman's system determines that the meaning of the user utterance "I would like to open this" is a request to perform an "open" command, which is determined by matching the recognized words and phrases to an "open" command in an application's symbolic language associated with a context. (Schmandt, ¶120; Coffman, 6:32-7:3.)

Under the combination, Coffman's process of identifying an acceptable context by matching the words and phrases also discloses **"the one or more recognized words or phrases are further associated with the determined context in response to the one or more recognized words or phrases satisfying the predetermined confidence level."** This is because the words and phrases are used to match a context and its associated formulated request (e.g., in symbolic language or application call/query). (Schmandt, ¶121 (citing Coffman, 42:1-7, 37:2-4).) Therefore, as a result of identifying a context, the words and phrases are associated with that determined context. (*Id.*) Moreover, as explained for 1[b][iii], it would have been obvious in view of Kanevsky to perform the claimed **"associating"** only when the recognized words and phrases satisfied a predetermined confidence level. (Schmandt, ¶121.)

(d) “**formulating...**” (1[d])

1[d] recites a “**domain agent.**” The ’006 patent describes “agents” as “autonomous executables that receive, process and respond to user questions, queries and commands” and states that they “provide complete, convenient and re-distributable packages or modules of functionality, typically for a specific domain or application.” (’006, 2:56-65.) In the software context, an “agent” generally broadly refers to a program that performs actions on behalf of another entity, such as a user or another software program or computing system. (Schmandt, ¶122 (examples of web browser as “HTTP user agent” and desktop email software as “mail user agent” or MUA).)

A “**domain agent**” in Coffman corresponds to the executable software of a conversationally-aware application (and associated data, such as vocabulary, grammar, language models and NLU symbolic mapping). (Schmandt, ¶123; Coffman, 13:26-29.) Coffman’s conversationally-aware application receives, processes, and responds to user questions, queries and commands. (Coffman, 13:8-13 (queries transformed into “application calls” and application output returned to user), 37:2-4 (task executes query), 40:9-19 (“execute completed queries” and “output of executed or running processes”).) Coffman’s conversationally-aware application can also provide functionality related to a specific domain, such as a particular subject matter (such as email or spreadsheet) or device (such as a PDA or

smart phone). (Coffman, 16:35-17:7 (listing “application specific” foundation classes, including “Manage\_e-mail,” “Manage\_calendar,” and “Manage\_addressbook/director”), 56:11-14 (e-mail), 58:30-32 (spreadsheet), 56:22-25 (“conversational PDA,” “conversational car PC (PVA, Personal Vehicle Assistant),” and “conversational smart phone”).) Coffman also describes modular development tools for conveniently building complete and re-distributable applications. (Coffman, 58:12-20; *see also id.*, 58:27-28, 7:29-8:9.) The applications can be distributed for installation by conversational systems. (Coffman, 52:9-11 & 52:18-21 (describing agents to upgrade or install applications).)

A conversationally-aware application executable is a “**domain agent**” because, as explained, it provides functionality for a particular area or realm, such as a particular subject matter or a device. (Schmandt, ¶121; Coffman, 16:35-17:7, 56:11-14, 58:30-32, 56:22-25, 36:29-31 (“[T]he present invention provides the capability to have a natural dialog with NLU, NLG and mixed initiative across multiple applications, with multiple domains.”).)

The ’006 patent’s description of “agents” further states that “[a]gents are autonomous executables.” (’006, 2:56.) A conversationally-aware application (“**domain agent**”) is an “autonomous executable” because it automatically (without user involvement) registers with a conversational system to be able to receive, process and respond to requests, and moreover, operates automatically in response

to the system selecting it to handle a request. (Schmandt, ¶125.) Coffman explains, for example, that “conversational devices or applications [] register their conversational capabilities” and describes “automatically discover[ing] local and network conversationally aware systems.” (Coffman, 7:12-14, 25:18-24; *see also id.*, 23:16-20.) Coffman also explains that “[t]he first task to have its dialog completed by producing a non-ambiguous query is executed” (Coffman, 37:2-4), making clear that an application is automatically selected and carries out a request as a result of processing by the conversational system. (*See also id.*, 42:1-7; Schmandt, ¶125.)

The autonomous nature of Coffman’s conversationally-aware applications is confirmed by **Maes (EX1007)**, which Coffman incorporates by reference and which claims priority to the same provisional patent applications as Coffman. (Coffman, 22:25-29; Coffman 1:2-3 & Maes, 1:4-5; EX1023 (Maes application corresponds to “YO999-113P”); Schmandt, ¶126.) Coffman specifically points to Maes for aspects of its conversational computing system. (Coffman, 22:25-29, 26:21-25, 45:11-22, 61:5-6.) Maes describes techniques for conversationally-aware applications to automatically communicate and coordinate with each other. (Maes, 3:4-10 (describing “automatically shar[ing] conversational resources automatically and in a coordinated and synchronized manner so as to provide a seamless conversational interface through an interface of one of the network devices.”).)

The preamble recites “**context-specific domain agents.**” 1[d] does not recite that the “**domain agent**” is “**context-specific.**” But assuming the preamble imposes a limitation, and that the “**domain agent**” in 1[d] must be “**context-specific,**” a conversationally-aware application in Coffman meets this requirement. As explained for 1[b][iii], Coffman describes a “context stack” where the state and history of each active process, such as an application process, is tracked in its own “context.” (Coffman, 29:18-30:3, 30:4-10.) Therefore, a conversationally-aware application is “**context-specific**” by virtue of its association with a context and lack of an association with other contexts. (Schmandt, ¶127.)

The prior art also discloses and renders obvious “**formulating, at the parser, the request contained in the utterance in accordance with a grammar used by a domain agent associated with the determined context.**” Coffman describes using natural language understanding to translate a recognized utterance into a symbolic mapping and query/call for an application to carry out. (Schmandt, ¶129; Coffman, 13:8-13 (“CVM 14 to transform queries into application calls”), 30:8-10 (“symbolic language (attribute value n-uple) which maps the actual query”).)<sup>4</sup> For example, Coffman teaches that a user’s utterance “I would like to open this” would be processed into “a set of attributes value pairs: Command: Open, Object:

---

<sup>4</sup> An “n-uple” refers to a collection of “n” elements; for example, a “4-uple” (or “quadruple”) is a collection of four elements. (Schmandt, ¶113 n.6.)

Window” and result in an “open” command being performed. (Schmandt, ¶129 (citing Coffman, 6:32-7:3).) And Coffman, as noted, explains that an application provides its active vocabulary, grammar, and language models to the CVM for controlling the application. (Coffman, 24:31-25:9, 57:22-27, 36:31-32.)

A POSITA would have understood and found it obvious that, in a system such as Coffman’s involving speech recognition using language model(s) and natural language understanding, a grammar associated with an application would be used to formulate the request as a symbolic mapping (such as a set of attribute-value pairs such as “Command: Open, Object: Window”) representing the system’s understanding and to, in turn, formulate an application query/call in accordance with that understanding. (Schmandt, ¶¶130-131 (citing Coffman, 53:25-26, 29:1-4, 6:32-7:3; EX1008, pp.0035 (“Natural Language Processing”), 0040 (“parsing”)).) Both the symbolic mapping and the application query/call would separately correspond to the claimed **“request contained in the utterance” “formulat[ed]...in accordance with a grammar used by a domain agent associated with the determined context.”** (*Id.*)

**Zadrozny** (EX1006) confirms the obviousness of formulating a request **“in accordance with a grammar used by”** a conversationally-aware application (**“a domain agent”**) associated with a context that was determined to be the best match for the user’s utterance. (Schmandt, ¶132.) **Zadrozny** is referenced in Kanevsky as

an example of a semantic analyzer. (Kanevsky, 14:22-25.) Zadrozny describes a “grammar of constructions” associated with an application (such as a calendar application) for performing natural language understanding to formulate a recognized utterance as an intended action and parameters and then translate the action and parameters into “a form that can be executed by the application.” (Zadrozny, pp.0019-0020.) Zadrozny illustrates the technique with the example utterance “I want you to arrange a conference in my office at 5.” (Zadrozny, p.0021.) Using a grammar of constructions, the utterance is “analyzed as *sent(assert, svoc)*, an assertive sentence consisting of a subject, verb, an object and a complement” and formulated as an action (“schedule”) and series of parameters:

```
[ [ action_name schedule]
  [ event_name
    [ a conference]
  [ event_time
    [ [ minute 0]
      [ hour
        [ 5 am_or_pm]
      ]
    ]
  [ event_place
    [ my office]
```

(Zadrozny, pp.0021-0022.) This action and parameters are then translated into a form for the application to execute (i.e., an application call/query).

Zadrozny’s process is just like Coffman’s descriptions of formulating a user’s utterance as a set of attribute-value pairs (e.g., “Command: Open, Object: Window”)

and then converting them into an application call/query. (Schmandt, ¶133; *see also* Coffman, 30:8-10 (“symbolic language (attribute value n-uple) which maps the actual query”), 13:8-13 (CVM system transforms queries into “application calls”).)

**Rationale and Motivation to Combine (Coffman, Kanevsky and Kneser with Zadrozny).** It would have been obvious to combine Zadrozny’s teachings regarding processing speech using a grammar of constructions with the conversational computing system taught by the combination of Coffman, Kanevsky and Kneser. (Schmandt, ¶134.) The combination would have predictably resulted in the conversational system in which a user request was formulated in accordance with a grammar used by a conversationally-aware application associated with the context determined to be the best match for the user’s request. (*Id.*)

Zadrozny is an analogous reference in the field of the ’006 patent, and would have been reasonably pertinent to problems facing the ’006 inventors, including techniques for understanding natural language speech input and how to determine and formulate actions for an application to undertake in response to the input. (Schmandt, ¶135.)

A POSITA would have been motivated to combine. Kanevsky expressly cites Zadrozny as an example of a “semantic analyzer.” (Kanevsky, 14:22-25.) And authors of Zadrozny are identified as being associated with IBM, the assignee/applicant for both Coffman and Kanevsky. (Zadrozny, p.0019 (“IBM

Research Division”).) A POSITA therefore would have turned to Zadrozny for detail about how to implement the systems and techniques described in Coffman and Kanevsky. (Schmandt, ¶136.)

A POSITA would have appreciated that the disclosures of Coffman and Zadrozny confirm their fit. (Schmandt, ¶137.) Coffman mostly describes the natural language understanding process at a high level: applications provide grammars, vocabularies, language models and symbolic mappings, a context is matched by “fit[ting] the utterance at the level of the NLU symbolic language” and “CVM 14 [] transform[s] queries into application calls.” Because Kanevsky expressly cites Zadrozny as an example of a semantic analyzer, a POSITA would have turned to it for its teachings regarding the use of grammars in natural language understanding to translate a recognized utterance into an application call. (*Id.*) Moreover, as noted, Zadrozny’s process matches Coffman’s descriptions of formulating a user’s utterance as a set of attribute-value pairs and then converting them into an application call/query. (*Id.*)

A POSITA would have had a reasonable expectation of success. (Schmandt, ¶138 (describing Zadrozny’s working implementation).)

**(e) “processing...” (1[e])**

As explained, Coffman teaches that the conversationally-aware application associated with the determined context (“**the domain agent associated with the**

**determined context**") is invoked to process the user's request. (Coffman, 37:2-4, 13:8-13 & 58:9-11, 40:9-19, 30:4-10.) Under the combination, as explained for 1[d], the conversationally-aware application processes "**the formulated request**" (the symbolic mapping and/or application call/query) because the application carries out processing in accordance with that request. (Schmandt, ¶139.) In Zadrozny's conference scheduling example, the calendar application would carry out processing to schedule a conference in the system in accordance with the specified action and parameters. (*Id.*)

A POSITA would have further understood and found it obvious that the "**processing**" is performed "**to generate a response to the utterance.**" (Schmandt, ¶140.) For example, Coffman describes executing completed queries and "return[ing] output of executed or running processes" and presenting them to user using, for example, text-to-speech, as Coffman illustrates in Figure 6:

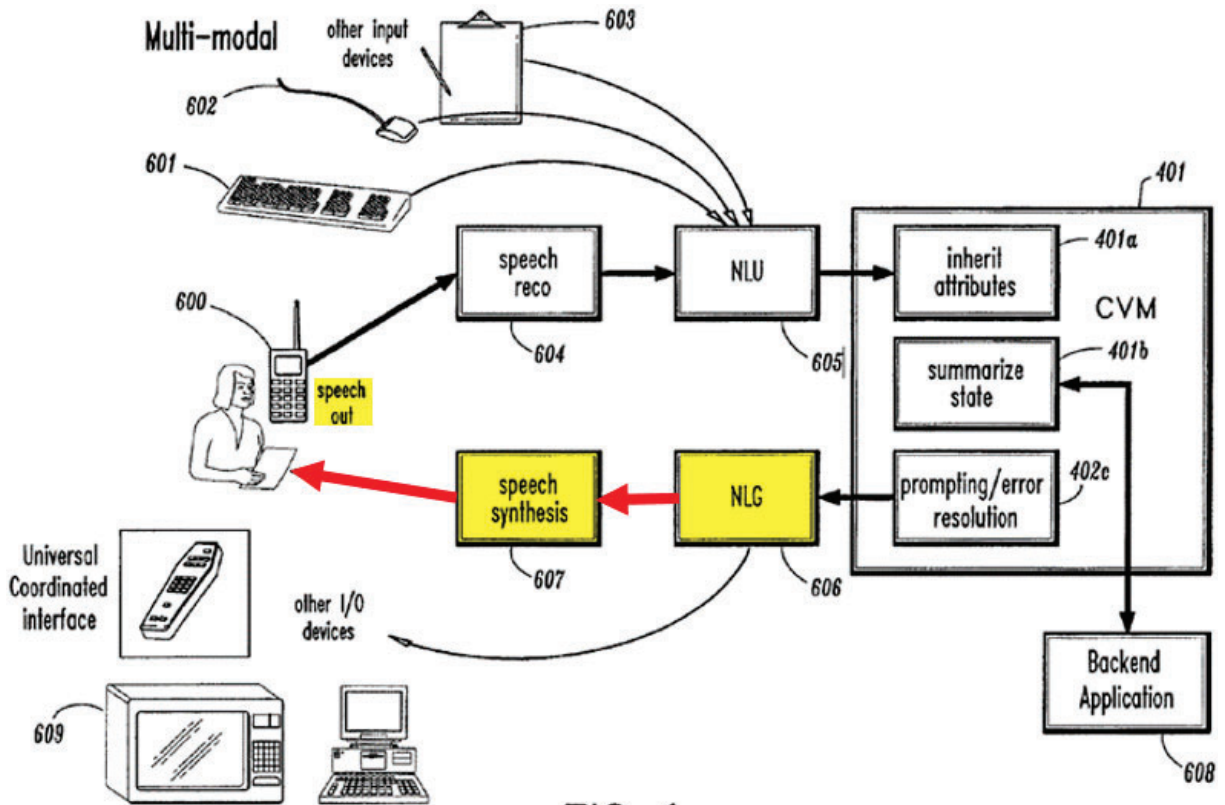
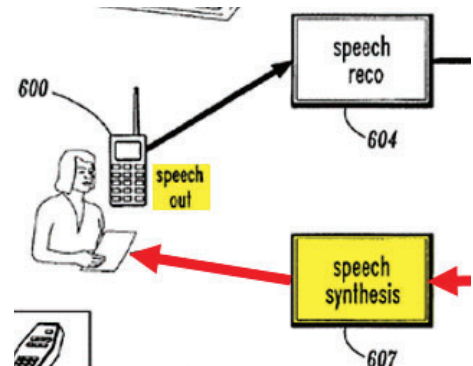


FIG. 6

(Coffman, Fig. 6 (annotations added), 40:9-19; *see also id.*, 13:8-13 & 58:9-11, 39:3-5.)

(f) “presenting...” (1[f])

Coffman discloses presenting to the user results generated by the conversationally-aware application (“**presenting the generated response to the utterance**”) as a result of processing the symbolic mapping and/or application call/query. (Schmandt, ¶141 (citing Coffman, Fig. 6, 40:9-19, 13:8-13, 58:9-11, 39:3-5).) Coffman further makes clear that the response can be presented “**via the speech unit,**” explaining that the



system can use “text-to-speech” to output the response. (Coffman, Fig. 6 (excerpt at right, showing speech synthesis 607 and “speech out” at telephone 600), 13:8-13 & 58:9-11 (CVM converts application “output into speech...provided to the user”).)

## 2. Claim 2

It would have been obvious for “**parsing the one or more recognized words or phrases to determine the context associated with the request**” to include the following steps:

### (a) “**matching...**” (2[a])

A POSITA would have understood and found it obvious that in Coffman’s system, words in the recognized speech (“**the one or more recognized words or phrases**”) would be “**matched**” to a “**set of keywords**”—words in, for example, an active vocabulary, grammar or NLU symbolic language provided by a conversationally-aware application—to determine the context corresponding to the user’s query. (Schmandt, ¶143.) Using Coffman’s example, the word “open” in the request “I would like to open this” matches an “open” command in an application’s symbolic language associated with a context. (Coffman, 6:30-7:3.) More generally, Coffman describes “match[ing]” recognized words and phrases against an NLU symbolic language and “mapping” user input into a symbolic language. (Coffman, 42:1-7; *see also id.*, 53:22-26.) Kanevsky, as noted, similarly teaches performing

natural language understanding “preferably” with a “combination of classical parsing capabilities with key word spotting.” (Kanevsky, 14:1-4.)

With respect to the “**set of keywords**” being “**associated with the dynamic set of prior probabilities or fuzzy possibilities**,” under the combination, the keywords would match words recognized as a result of applying adaptive language model probabilities (“**the dynamic set of prior probabilities**”)—for example, the keyword “open” in a symbolic language matches the recognized word “open.” As a result, the keywords would also be “**associated with the dynamic set of prior probabilities or fuzzy possibilities**.” (Schmandt, ¶144.)

(b) “scoring...” (2[b])

Coffman discloses determining acceptability of “**multiple possible contexts for the one or more recognized words or phrases matched to the set of keywords using...a context stack that includes one or more recent contexts**.” (Schmandt, ¶145.) Coffman describes searching a context stack to find a context that is “an acceptable match” by “fit[ting] the utterance at the level of the NLU symbolic language.” (Coffman, 42:1-5.) If the match with a context in the context stack is “too improbable,” then the next context in the context stack is considered for matching, thereby “**using...a context stack**.” (Coffman, 40:24-27.) Coffman also makes clear that its context stack “**includes one or more recent contexts**.” (Coffman, 29:20-30, 30:13-16, 41:8-9; Schmandt, ¶145.)

A POSITA would have further understood and found it obvious that Coffman's technique of determining whether a context was an "acceptable match" or "too improbable" would involve "**scoring**" (a well-known concept) the context, such as determining the likelihood or goodness of the match in order to compare it against some level of acceptability. (Schmandt, ¶146.) Kanevsky confirms the obviousness of "**scoring**" a context, by expressly describing multiple analogous situations (speaker recognition and semantic understanding) in which a score is generated (and compared against a predetermined threshold value) to determine further action to take. For example, Kanevsky describes a process for recognizing the speaker that involves both scoring a user's spoken input against a stored answer and scoring a user's voice sample against a stored user model. (Kanevsky, 6:34-42, 7:12-20.) A combined score is then compared "to a predetermined threshold value and if the score is one of substantially equivalent to and above the threshold value," the speaker is permitted access to the system. (Kanevsky, 3:40-48; *see also id.*, 7:12-20.)

It would have been obvious to apply these scoring/matching techniques to Coffman's context matching. (Schmandt, ¶147.) The rationale and motivation to combine Coffman and Kanevsky discussed above applies here. (*Id.*) Specifically, Coffman describes rejecting a context where a match is "too improbable," which confirms that there is some level of "probability" required. Kanevsky confirms that

using a scoring technique would have been a straightforward and obvious choice—applying Kanevsky’s known “scoring” techniques to improve Coffman’s methods for determining a context matching a user’s query amounts to improving a similar method in the same way, with predictable results. (*Id.*)

**(c) “selecting...” (2[c])**

As explained for 2[b], a POSITA would have understood and found it obvious that the process of searching a context stack to select an acceptable context would involve scoring each searched context at least until a context having a score high enough to satisfy a predetermined threshold value was identified. (Schmandt, ¶148 (citing Coffman, 40:24-27, 42:1-5, 41:8-11).) The context having such a high score would then be **“selected...to be the determined context,”** and the application associated with that context would process the user’s request. (*Id.*)

**3. Claim 3**

It would have been obvious for **“parsing the one or more recognized words or phrases to determine the context associated with the request”** to further include the following steps:

**(a) “requesting...” (3[a])**

Coffman’s system will prompt the user to verify their input as the system understood it (**“recognized words and phrases”**) if the system determines that the input was incomplete or ambiguous—i.e., context scores do not satisfy a

predetermined confidence level. (Schmandt, ¶150; Coffman, 41:16-21 (“generate a request (a prompt) for the missing or ambiguous information and update the context (requested missing fields)”); *see also* Coffman, 37:2-4, 40:9-19 (“provide feedback to the user for incomplete or ambiguous queries/command”).)

**(b) “determining...” (3[b])**

Because Coffman describes prompting a user to provide missing or ambiguous information, resulting in an updated/completed context, a POSITA would have understood and found it obvious for Coffman’s system to **“determine the context based on [the additional] information provided by the user.”** (Schmandt, ¶151.) A POSITA would have understood and found it obvious for the user to provide the additional information via **“one or more subsequent natural language speech utterances.”** (*Id.* (citing Coffman, 14:27 (“spoken prompt”), 5:18 (“speech input”))).

**C. Ground 2: Claim 4 Is Obvious Over Ground 1 Prior Art in Further View of Maes ’101 and Franceschi.**

**1. Claim 4**

As shown below, the claimed method is obvious over Ground 1 prior art in further view of **Maes ’101 (EX1010)** and **Franceschi (EX1011)**.

**(a) “receiving...” (4[a])**

4[a] and 1[a] are identical; for the reasons explained for 1[a], 4[a] is obvious. (Schmandt, ¶153.)

**(b) “recognizing...” (4[b])**

4[b] and 1[b] are identical; for the reasons explained for 1[b], 4[b] is obvious. (Schmandt, ¶154.) As explained below, the prior art also discloses and renders obvious that **“recognizing the one or more words or phrases contained in the utterance includes”** the following steps.

**(i) “requesting...” (4[b][i])**

As explained, Coffman’s system will **“request[] a user that spoke the utterance”** to verify their input as the system understood it if the system determines that input was incomplete or ambiguous. (Coffman, 41:16-21, 40:9-19, 37:2-4; Schmandt, ¶156.) Coffman does not expressly describe prompting the user to **“spell the one or more words or phrases in response to the one or more recognized words or phrases failing to satisfy a predetermined confidence level.”** But as explained below, applying spelling to Coffman’s techniques for prompting a user regarding ambiguous and/or incomplete information would have been obvious. (Schmandt, ¶157.)

For starters, the use of a predetermined confidence level to accept or reject recognition results was a routine and conventional feature of speech recognition systems, and would have formed part of a POSITA’s background knowledge. (Schmandt, ¶158 (citing EX1008, pp.0026 (“confidence measure”), 0046 (“rejection”); EX1017, pp.002-003 (describing “degree of certainty for recognition

of each word”)).) Moreover, Coffman incorporates Maes (EX1007) (Coffman, 22:25-29), which acknowledges “efforts...to develop appropriate confidence measures for [] recognized speech” (Maes, 17:23-26) and describes rejecting speech recognition results “based on confidence measures” and determining whether speech recognition results meet “a given threshold.” (Maes, 16:20-17:7 (“the rejection of results for speech recognition and NLU may occur...based on confidence measures....[A]ny suitable measure or approach to estimate the confidence or the recognized utterance or query (during or after the dialog) can be used to determine if the results obtained by a conversational system are acceptable or not”), 17:23-18:16.)

Rather than outright rejecting the user’s utterance when speech recognition results fall below a given confidence threshold and, for example, asking the user to repeat themselves, it would have been obvious to prompt the user to spell one or more words. (Schmandt, ¶160.) This would further the goal of an adaptive language model to be able to better recognize a user’s utterances. (*Id.*) Voice spelling would also provide a convenient way to provide input in situations where other forms of input were limited or less desirable. (*Id.*) The obviousness of prompting a user to spell word(s) is confirmed by another patent to Stephane Maes, Maes ’101 (EX1010), which explains that “unrecognized utterances can be defined by the user during enrollment or later enrollment correction by prompting the user to spell the

word whenever speech recognition is unsuccessful.” (Maes ’101, 6:52-55; Schmandt, ¶160.) Maes ’101 discloses that, by spelling words, the system’s performance can be improved and adapted to the user by adding the spelled word to the speech recognizer’s vocabulary. (Maes ’101, 6:55-66.) 4[b][i] therefore would have been obvious in further view of Maes ’101. (Schmandt, ¶160.)

**“Phonetic alphabet spelling” Issue:** While 4[b][i] simply recites “requesting a user...to spell,” 4[b][ii] refers to “a phonetic alphabet spelling” provided by the user. It would have been obvious for the user to provide a “phonetic alphabet spelling” of the one or more words or phrases. (Schmandt, ¶161.) This is confirmed by Franceschi (EX1011), which explains motivations and benefits of phonetic alphabet spelling, and that such functionality was routine and conventional, with commercially-available speech recognition systems already including it. (Franceschi, 1:38-52, 2:8-13.)

**Rationale and Motivation to Combine (Ground 1 Prior Art with Maes ’101 and Franceschi)**. It would have been obvious to further combine with Maes ’101 and Franceschi. The combination would have predictably resulted in Coffman’s conversational system with the ability to prompt the user to provide a phonetic alphabet spelling if speech recognition results failed to meet a certain confidence level. (Schmandt, ¶162.)

Maes '101 and Franceschi are both analogous references in the field of the '006 patent, and would have been reasonably pertinent to a number of problems facing the patent inventors, including techniques for receiving speech input from a user and improving the reliability of speech recognition results by resolving ambiguity. (Schmandt, ¶163 (citing Maes '101, 1:11-18; Franceschi, 1:6-8).)

A POSITA would have been motivated to combine. Coffman and Maes '101 (and the Maes reference incorporated into Coffman) name a common co-inventor, Stephane Maes, and identify the same assignee/applicant, IBM. (Coffman, face page; Maes '101, face page; Maes, face page.) Franceschi likewise identifies IBM as the assignee. (Franceschi, face page.) Given the connections between these references, therefore, a POSITA would have appreciated that Maes '101 and Franceschi could be considered in arriving at an understanding of Coffman's conversational system. (Schmandt, ¶164.)

Moreover, as noted, Coffman teaches that the system prompts the user to verify their input if the system determines that the input was incomplete or ambiguous. (Coffman, 41:16-21, 37:2-4, 40:9-19.) Franceschi teaches that “it frequently occurs that the best approach for an unrecognized, difficult, or new word is for the speaker to spell it.” (Franceschi, 1:38-40.) A POSITA would have therefore appreciated that this “frequently best approach”—prompting a user to spell a word—would help resolve incomplete or ambiguous input, and that the ability to

provide a phonetic alphabet spelling as taught by Franceschi would help avoid ambiguity when processing the user's spelling input, resulting in an improved user experience. (Schmandt, ¶165.)

A POSITA would have had a reasonable expectation of success, as it would have involved conventional techniques and technologies without undue experimentation. (Schmandt, ¶166 (citing Franceschi, 2:8-13).)

(ii) “updating...” (4[b][ii])

As noted for 4[b][i], Maes '101 discloses that, by spelling words, the system's performance can be improved and adapted to the user, by adding the spelled word to the speech recognizer's vocabulary. (Maes '101, 6:55-66.) Adding words to a speech recognizer's vocabulary would also further Kneser's goal of improving a speech recognizer's language model, “[s]ince the quality of the adaptation of the language model values has a considerable influence on the reliability of the speech recognition.” (Kneser, 1:25-29.) It would have been obvious, therefore, to update trigram language model(s) used for speech recognition (“**the one or more dictionary and phrase tables**”) to incorporate information about the additional vocabulary provided via phonetic alphabet spelling (“**updat[e] the information in the one or more dictionary and phrase tables based on a phonetic alphabet spelling associated with the one or more words or phrases**”). (Schmandt, ¶168.)

Under the combination, and as Maes '101 teaches, the user would provide the spelling by voice after being prompted (“**the user provides the phonetic alphabet spelling in one or more subsequent natural language speech utterances**”). (Maes '101, 6:52-66; Schmandt, ¶169.)

(c) “**parsing...**” (4[c])

4[c] is substantially similar to 1[f], and the analysis of 1[f] applies. (Schmandt, ¶170.) 4[c] further includes “**parsing...the phonetic alphabet spelling associated with the one or more words or phrases.**” A POSITA would have understood and found it obvious that the user’s phonetic alphabet spelling would be parsed under the combination with Coffman. (*Id.*) Parsing the spelling (such as parsing “Juliet Oscar Echo” to determine the user was spelling “Joe”) would have enabled the Coffman’s system to determine the spelled word so that the system could include the word as part of determining the meaning of the user’s utterance and an associated context. (*Id.*)

(d) “**formulating...**” (4[d])

(e) “**processing...**” (4[e])

(f) “**presenting...**” (4[f])

4[d]-[f] are identical to 1[d]-[f]. For the reasons explained for 1[d]-[f], 4[d]-[f] are obvious. (Schmandt, ¶¶171-173.)

**D. Ground 3: Claim 5 Is Obvious Over Ground 1 Prior Art (Coffman, Kanevsky, Kneser and Zadrozny) in Further View of Chai.**

**1. Claim 5**

As shown below, the claimed method is obvious over Ground 1 prior art in further view of Chai (EX1012).

(a) “receiving...” (5[a])

(b) “recognizing...” (5[b])

5[a] and 5[b] are identical to 1[a] and 1[b], except that 5[b] does not further recite “wherein recognizing the one or more words or phrases contained in the utterance includes.” Therefore, for the reasons explained for 1[a] and 1[b], 5[a] and 5[b] are obvious. (Schmandt, ¶¶175-176.)

(c) “parsing...” (5[c])

For purposes of this Petition, 5[c] is substantially identical to 1[c]. (Schmandt, ¶177.) For the reasons explained for 1[c], 5[c] is obvious. (*Id.*)

(d)

(e) “formulating...” (5[d])

5[d], like 1[d], recites “**formulating, at the parser, the request contained in the utterance in accordance with a grammar used by a domain agent associated with the determined context.**” This portion of 5[d] is therefore disclosed by and obvious over the combination. (Schmandt, ¶178.)

5[d] further recites “**wherein formulating the request in accordance with the grammar used by the domain agent includes**” the following steps. These steps would have been obvious in further view of Chai (EX1012).

(i) “**determining...**” (5[d][i])

As discussed for 1[d], a POSITA would have understood that, in a system such as Coffman’s involving speech recognition using language model(s) and natural language understanding, a grammar associated with an application would be used to understand the recognized speech so that a symbolic mapping and application query/call could be formulated in accordance with the understood meaning. (Schmandt, ¶180.) This is confirmed by Zadrozny, which as discussed for 1[d], discloses a “grammar of constructions” that is likewise used to determine an intended action and its parameters for an application to execute from a user’s natural language input. (Zadrozny, pp.0019-0020.) Therefore, under the combination, where the application would be a conversationally-aware application of Coffman (“**the domain agent**”), the determined action and parameters for criteria using Zadrozny’s “grammar of constructions” would correspond to “**the request in the grammar used by the domain agent.**” (Schmandt, ¶180.)

For example, Zadrozny describes using a grammar of constructions to translate the statement “I want you to arrange a conference in my office at 5” into an

action (“action\_name = schedule”) with various criteria and parameters (e.g., “event\_name [ a conference ]”; “event\_place [ my office ]”):

For example, the sentence *I want you to arrange a conference in my office at 5* is analyzed as *sent(assert, svoc)*, an assertive sentence consisting of a subject, a verb, an object and a complement.

...

This result of parsing is then interpreted by the domain interpreter to produce:

```
***Slots:
[ [ action_name schedule]
  [ event_name
    [ a conference]
  [ event_time
    [ [ minute 0]
      [ hour
        [ 5 am_or_pm]
    [ event_place
      [ my office]
```

(Zadrozny, pp.0021-0022.)

As shown above, Zadrozny identifies the various fields for actions and criteria/parameters as “slots” (“\*\*\*Slots:”).<sup>5</sup> Zadrozny’s technique of mapping the words and phrases of the parsed input to various slots related to a grammar of constructions is consistent with Coffman’s descriptions of determining a context and query by filling in symbolic fields for a natural language query. (Schmandt, ¶182; Coffman, 38:15-16, 42:4-5, 30:8-10.)

---

<sup>5</sup> As Petitioner’s expert explains, the “slots” technique was well known in the art. (Schmandt, ¶182 n.5.)

Zadrozny does not expressly disclose that an action based on a grammar of constructions (“**the request in the grammar used by the domain agent**”) would have “**one or more required values and one or more optional values**” but this would have been obvious over Chai.

Like Coffman and Zadrozny, Chai describes techniques for processing a user’s natural language request to determine an application action to execute. Like the symbolic field filling described in Coffman and the slot filling disclosed in Zadrozny, Chai describes filling “slots” in “action templates” based on natural language input to determine the application action to execute. (Chai, 4:51-5:11 (“A natural language parser parses one or more natural language phrases...into one or more logical forms (parsed user input)....[M]odified logical forms are matched against a set of predefined action templates, the best matching template is selected, corresponding action requests are sent to an action manager...”).)

For instance, in an online shopping example, Chai describes processing two user requests, “I want to buy one 30 inch wide White Color Cooktop with Electric PowerSource and with Electric:Radiant Glass CookTop Surface and of KenMore company” followed by “I want to buy one Black CookTop of previous type.” (Chai, 12:28-32.) Chai explains that the user’s input is used to fill various slots of matched action templates, with some slots being “**required**”—for example, a value for the “item name” slot is required:

Suppose the user input **205** was “I want to buy one 30 inch wide White Color Cooktop with Electric PowerSource and with Electric:Radiant Glass CookTop Surface and of Ken-More company” followed by another request “I want to buy one Black CookTop of previous type” Before taking discourse context into consideration, the following slots **807** in matching templates **806** with “buy” transaction are filled.

First Template (Used for Pants and Shirts):

TRANSACTION: buy

TEMPLATE SLOTS:

item name: CookTop

sku:

quantity: 1

size:

color: Black

company:

REQUIRED SLOTS: item name, quantity, company

(Chai, 12:32-44 (highlighting added); *see also id.*, 12:45-56 (second template), 12:21-22 (“choosing the template with the most number of required slots filled”).)

On the other hand, referring to the example action template above, because a value for the “color” slot is not identified as being required, a POSITA would have understood it to be optional. Chai therefore discloses both “**one or more required values**” (e.g., the value for the “item name” slot in Chai’s example) and “**one or more optional values**” (e.g., the value for the “color” slot in Chai’s example). (Schmandt, ¶185.)

### **Rationale and Motivation to Combine (Ground 1 Prior Art with Chai).**

It would have been obvious to combine Chai’s teachings regarding required and optional slot values for action templates with Zadrozny’s technique for formulating

a request using a grammar of constructions. (Schmandt, ¶186.) This would have predictably resulted in Coffman system’s step of determining the call/query for the conversationally-aware application (“**the domain agent**”) to include the step of determining required and optional values for the application call/query (“**determining one or more required values and one or more optional values associated with formulating the request in the grammar used by the domain agent**”). (*Id.*)

Chai is an analogous reference in the field of the ’006 patent, and would have also been reasonably pertinent to a number of problems facing the patent inventors, including how to formulate application queries using natural language. (Schmandt, ¶187 (citing Chai, 1:8-10).)

A POSITA would have been motivated to combine. Chai lists Wlodek Zadrozny, lead author of Zadrozny (**EX1006**), as a co-inventor and also identifies IBM as the assignee. (Chai, face page.) A POSITA would have appreciated the relevance of Chai for understanding Zadrozny’s techniques for formulating application queries and Coffman’s conversational system. (Schmandt, ¶188.) While Chai focuses on online shopping examples, Chai also makes clear that its techniques are applicable in other situations similar to those discussed in Coffman, such as retrieving stock quotes. (Chai, 5:10 (“Retrieving the latest price of a stock”), 8:16-65 (describing “stock trading application”); Coffman, 42:50-52 (“check the news or

stock quotes”).) A POSITA would have therefore appreciated the broad applicability of Chai’s techniques to formulating queries for conversationally-aware applications in Coffman’s system. (Schmandt, ¶189.) A POSITA would have also appreciated that the ability to formulate queries with required and optional values would have benefitted Coffman’s system by providing a more flexible system for responding to user utterances. For example, the ability to provide additional, but not strictly necessary, information as part of a request would enable a user to more quickly and directly locate desired information rather than receive less precise information or engage in a lengthier colloquy with the system to refine the response to the user’s request. (*Id.*)

A POSITA would have had a reasonable expectation of success, as it would have entailed well known and conventional computer programming techniques without undue experimentation. (Schmandt, ¶190.)

(ii) “extracting...” (5[d][ii])

The ’006 patent generally describes “**criteria**” as attributes of a request, and “**parameters**” as the values of the attributes (although they are also described somewhat interchangeably). (Schmandt, ¶191.) For example, for a request in a weather context, a criterion could be location, and the parameter would be the specific value for the location (e.g., “Alexandria”). (’006, 18:30-32; *see also id.*, 25:15-18 (same), 11:65-67.)

It would have been obvious to “**extract[] one or more criteria and one or more parameters from one or more keywords contained in the one or more recognized words or phrases.**” As explained, Coffman describes determining a context by “fit[ting] the utterance at the level of the NLU symbolic language” (Coffman, 42:4-5) and filling symbolic fields for a query (Coffman, 38:15-16, 30:8-10, 41:16-21, 6:32-7:3). Coffman does not provide an example of how exactly an utterance is “fit...at the level of the NLU symbolic language” or how “symbolic fields are filled,” but it would have been obvious under the combination with Zadrozny and Chai to extract criteria and parameters from words contained in the recognized utterance. (Schmandt, ¶192.) For example, as explained, Chai describes analyzing a user’s input (“I want to buy one Black CookTop of previous type”) to identify various slots (“**one or more criteria**”) (e.g., color) and their values (“**one or more parameters**”) (e.g., black) for an action template.

It would have also been obvious for the claimed “**extracting**” to “**us[e] procedures sensitive to the determined context.**” As explained, Coffman describes using context to determine a conversationally-aware application and call/query. Therefore, by matching a user’s words and phrases to fields in the symbolic language of an application call/query, Coffman’s techniques, under the combination, would have predictably resulted in “**extracting one or more criteria**

**and one or more parameters...using procedures sensitive to the determined context.**” (Schmandt, ¶193.)

**(iii) “inferring...” (5[d][iii])**

Chai discloses filling “slots” in action templates with values based on words and phrases input by a user, such as the user’s statement, “I want to buy one Black CookTop of previous type.” (Chai, 12:28-32.) The user’s statement, however, does not fill all required slots in the action template(s)—e.g., a value for the required “company” slot is missing. (Chai, 12:28-56.) Chai teaches that missing slots can be filled in by “**inferring**” values from context. (Chai, 12:1-20 (“...determine if any of the missing slot **807** values can be inferred from context...”).) Chai therefore discloses “**inferring one or more further criteria and one or more further parameters**” by reviewing context to infer that a value (“**parameter**”) corresponds to a particular slot (“**criteria**”) whose value is missing from an action template. (Schmandt, ¶¶194-195.)

It would have been obvious to apply Chai’s techniques to Coffman. (Schmandt, ¶196.) As noted, Coffman discloses a “context” for an active process that includes similar information to Chai, including “past history of the discourse, its current state, and the specific characteristics and full query arguments” and “any data needed for input understanding.” (Coffman, 29:18-30:3.) A POSITA would have therefore readily appreciated that Chai’s techniques to “infer” criteria and

parameters could be applied to Coffman's system, and recognized the similar benefits that would flow from it, including the ability to complete a user's query without prompting the user to supply information that the system could determine on its own. (Schmandt, ¶196.) It therefore would have been obvious that recognized speech using trigram probabilities ("**using a dynamic set of prior probabilities or fuzzy possibilities**") from an utterance could provide context information used to fill in missing slot values, predictably resulting in the "**inferring**" step taking place by "**using a dynamic set of prior probabilities or fuzzy possibilities.**" (Schmandt, ¶197.)

(iv) "**transforming...**" (5[d][iv])

5[d][iv] corresponds to the step of filling in required and optional query fields (or action "slots") based on a grammar of constructions as taught by the combination of Coffman, Zadrozny and Chai. (Schmandt, ¶198.) "[T]**he one or more extracted criteria, the one or more extracted parameters, the one or more inferred criteria, and the one or more inferred parameters,**" as discussed for 5[d][ii] and 5[d][iii], correspond to recognized words and phrases from the user that are used to fill required and optional slots and their values. The recognized information under the combination would have also been "**transformed...into one or more tokens.**" This is because, consistent with the ordinary understanding of "**token**" as generally referring to an individuated or non-reducible text element in data being parsed, the

recognized information would be used to construct a request having a format consistent with the grammar of constructions and its slots as taught by Zadrozny, which would in turn be formatted into a call/query for the conversationally-aware application to process. (Schmandt, ¶198 (citing EX1009, p.0035 (“token”)).) And both of these formats would have been requests formulated **“in the grammar used by the domain agent”**—the initial formulation because it would use the grammar of constructions to identify and fill in slots, and the second formulation (for an application call/query) because it is based on and derived from that initial formulation. (*Id.*) By using the tokens to identify and fill in slots, moreover, the tokens would **“hav[e] a format compatible with the grammar used by the domain agent.”** (*Id.*)

It would have been obvious for **“the one or more tokens [to] include all the required values and one or more of the optional values”** for the formulated requests. (Schmandt, ¶199.) As noted, Coffman describes executing “completed” queries, while providing feedback for “incomplete or ambiguous queries/commands.” (Coffman, 40:9-19.) Chai teaches that a “complete” query can involve a query where all required slots are filled and some or all of the optional fields are filled. (Chai, e.g., 13:15-19.)

(f) “processing...” (5[e])

(g) “presenting...” (5[f])

5[e] and 5[f] are identical to 1[e] and 1[f]. Therefore, for the reasons explained for 1[e] and 1[f], 5[e] and 5[f] are obvious over the combination. (Schmandt, ¶¶200-201.)

**E. Ground 4: Ground 4: Claims 6 and 9 Are Obvious Over Ground 1 Prior Art (Coffman, Kanevsky, Kneser and Zadrozny) in Further View of Redfern.**

**1. Claim 6**

As shown below, the claimed method is obvious over Ground 1 prior art in further view of **Redfern (EX1013)**. Redfern describes a meta-search application where a user can provide a single natural language query to the application, and the application will formulate queries to multiple search engines and return what the application deems the best search results by evaluating the results using the natural language query. (Redfern, Abstract.) As explained below, under the combination with Redfern, a meta-search application configured in accordance with Coffman to be conversationally-aware would be a “**context-specific domain agent.**” (Schmandt, ¶202.)

(a) “receiving...” (6[a])

As explained for 1[a], Ground 1 prior art discloses and renders obvious “**receiving, at a speech unit coupled to a processing device, a natural language speech utterance that contains a request.**” As explained below, it would have

further been obvious in view of Redfern for the “**request**” to “**include[] at least one command and at least one question.**”

With respect to a request that “**includes at least one command and at least one question,**” the ’006 patent states that a question and a command are similar but that a question typically results in a response while a command results in an action, and that an utterance can include “both aspects of a question and a command.” (’006, 5:59-62, 10:67-11:4, 3:46-52.)

It would have been obvious for the request in the user’s utterance to “**include at least one command and at least one question.**” Redfern provides an example of “**at least one question**”—“Where do Monarch butterflies spend the winter?”—which is translated into the request “rank1(Monarch) phrase(butterflies spend) phrase (winter)” to be processed by the meta-search application. (Redfern, 15:42-65.) It would have been obvious for the user’s question to be provided as part of a spoken utterance. (Schmandt, ¶205.)

The “**at least one command**” under the combination could have taken various forms. For example, the command could relate to selecting the set of search engines/information sources to query. Redfern explains that the command to specify the set of search engines/information sources to query could be explicit or implicit. (Redfern, 9:22-27.) It would have been obvious for a user to explicitly specify the set of search engines/information sources for a search with an utterance such as

“Search Lycos, AltaVista and Excite for where do Monarch butterflies spend the winter?” (Schmandt, ¶206.) As another example, the command for the set of search engines/information sources for a search could be implicitly specified by the particular keywords in the user’s query, consistent with the ’006 patent’s statement that an utterance can include aspects of both a question and a command. (’006, 3:46-52.) As another example, a well-known command during web browsing or other network-based activities is to stop/cancel a pending operation. It would have been obvious for a user’s utterance to include a command to stop/cancel a pending search operation in order to enter a new one. (Schmandt, ¶206.) These examples include a “**command**” because they instruct the system to take some action—such as select and use particular search engines or stop processing. The ability to enter commands of this nature using speech utterances would have been especially useful in situations and devices with limited input/output capabilities beyond voice as contemplated by Coffman. (*Id.*)

**Rationale and Motivation to Combine (Ground 1 Prior Art and Redfern).**

It would have been obvious to combine Redfern’s teachings regarding a meta-search application with Coffman’s teachings regarding conversationally-aware applications. (Schmandt, ¶207.) The combination would have predictably resulted in a meta-search application as taught by Redfern with the ability to interface with Coffman’s conversational system, thereby advantageously enabling, for example, a

user to use the meta-search application by providing spoken natural language input and with a variety of devices with different input and output capabilities. (*Id.*)

Redfern is an analogous reference in the field of the '006 patent, and would have been reasonably pertinent to problems facing the '006 inventors, including how to interface and control an application using natural language input and how to query multiple information sources using natural language input. (Schmandt, ¶208 (citing Redfern, 1:6-13).)

A POSITA would have been motivated to combine. (Schmandt, ¶209.) Coffman itself describes conversational search functionality as well as integration with websites. (Coffman, 48:25-49:16 (search capability), 58:32-59:3.) Redfern provides an express motivation for a natural-language-based meta-search, explaining that it is more efficient and less burdensome for the user. (Redfern, 1:66-2:8, 2:38-44.) Moreover, Redfern itself already describes the application's ability to receive natural language search data. (Redfern, 4:28-31.) A POSITA would have readily appreciated that adapting and incorporating Redfern's meta-search application for Coffman's conversational system would enhance a user's ability to input natural language search queries and instructions. (Schmandt, ¶209.)

A POSITA would have considered Redfern and Coffman to be combinable using conventional techniques and technologies. (Schmandt, ¶210 (citing Redfern, 4:28-31, 10:25-38, 4:8-9; Coffman, 12:5-7, 13:1-3, 22:32-23:3, 55:13-19).) A

POSITA would have therefore had a reasonable expectation of success, without undue experimentation. (Schmandt, ¶211.)

**(b) “parsing...” (6[b])**

For purposes of this Petition, 6[b] is essentially subsumed by 1[c]; therefore, for the reasons explained for 1[c], 6[b] is obvious over the combination. (Schmandt, ¶212.)

**(c) “formulating...” (6[c])**

6[c] is identical to 1[d], and therefore analysis for 1[d] applies. (Schmandt, ¶213.) Under the proposed further combination with Redfern, the recited “**domain agent**” would take the form of a meta-search application configured in accordance with Coffman to be conversationally-aware. (*Id.*) The meta-search application receives a natural language query and evaluates it to produce queries to submit to multiple information sources, such as the well-known Lycos, AltaVista and Excite web-based search engines. (Redfern, 2:52-3:9, 4:8-34, 8:66-9:12, 15:43-65.) The application then scores the results and extracts relevant information to present to the user. (Redfern, 2:52-3:9, 15:19-41, 16:47-60.)

Under the combination with Redfern, the formulated request would correspond to an application call/query to the meta-search application (“**domain agent**”) that included search data resulting from processing and formatting the user’s

natural language query. (Schmandt, ¶214; Redfern, 4:35-8:65 (describing processing and formatting), 15:44-49.)

**(d) “processing...” (6[d])**

6[d], like 1[e], recites “**processing the formulated request with the domain agent associated with the determined context to generate a response to the utterance**” and therefore this portion of 6[d] is disclosed by and obvious over the combination. (Schmandt, ¶215.) 6[d] further recites “**wherein processing the formulated request with the domain agent includes**” the following steps, which would have been obvious in further view of Redfern.

**(e) “directing...” (6[e])**

With respect to a “**command**” to select (explicitly or implicitly) the search engines/information sources to query, it also would have been obvious for the conversationally-aware meta-search application (“**domain agent**”) to process the application call/query (“**process the formulated request**”) by “**directing the at least one command**” to “**local or remote processing devices.**” For example, Redfern explains that, in order to select and use particular search engines, the application consults “a database of search engine capabilities, requirements and addresses (URL’s or other appropriate address information)...to determine the appropriate parameters for each search engine in the selected set of search engines.” (Redfern, 9:34-38.) By consulting such a database in this manner, the application

**“direct[s] the at least one command”** to the database system (**“local or remote processing devices”**). (Schmandt, ¶218.)

6[e] would have also been obvious for a **“command”** to cancel a pending search. For example, a POSITA would have understood that canceling a pending search operation would have entailed instructing local and remote processing devices to stop processing related to the search and close network connections. (Schmandt, ¶219.)

**(f) “querying...” (6[f])**

It would have been obvious under the combination for the conversationally-aware meta-search application (**“domain agent”**) to process the application call/query (**“process the formulated request”**) by **“querying one or more local or network information sources to resolve information requested in the at least one question”** (e.g., by sending search requests to multiple web-based search engines or other information sources). (Schmandt, ¶220; Redfern, 15:42-65, 10:25-33, 2:60-63.)

As explained below, it would have been obvious under the combination for **“querying the one or more local or network information sources [to] further include[]”** the following steps.

**(g) “submitting...” (6[g])**

As explained, Redfern’s meta-search application submits “multiple...queries” to “the one or more local or network information sources.” For example, to resolve the user’s question “Where do Monarch butterflies spend the winter?” the application submits a query of “Monarch+butterflies+spend+winter” to the Lycos search engine, a query (and further instruction to list results having “Monarch” in them first) of “Monarch+AND+(butterflies+OR+spend)+AND+(winter) Monarch ranked first” to the AltaVista search engine, and a query of “Monarch+AND+(butterflies+OR+spend)+AND+(winter)” to the Excite search engine. (Redfern, 15:42-65; *see also id.*, 25:19-21 (describing AltaVista “selection criteria” and “results ranking criteria”).)

Redfern further discloses that the submitted queries are “asynchronous” because the meta-search engine submits each query without waiting for an earlier one to be completed or time-out. (Redfern, 15:63-65, 10:38-44; *see also id.*, Fig. 5 (steps 528 and 532 showing that all queries are dispatched in step 528 and all results received at step 532); Schmandt, ¶223.) Moreover, it would have been obvious for the queries to be “asynchronous” in order to more quickly receive and return results to the user by having multiple search engines processing queries substantially simultaneously. (Schmandt, ¶223.) Redfern acknowledges the need for a timely

response to the user, describing obtaining results before “a predetermined time limit has been exceeded.” (Redfern, 10:38-44.)

As to “**the multiple queries include one or more duplicate queries submitted to distinct information sources**,” a POSITA would have understood and found it obvious that “**one or more duplicate queries**” would be submitted by the meta-search application across the selected set of search engines/information sources. (Schmandt, ¶224.) Nothing in the ’006 patent suggests that queries must be textually verbatim or in an identical format to be considered “**duplicate queries**.” Redfern’s example queries about Monarch butterflies are three “duplicate” queries because they are substantively identical in seeking the same information. (*Id.* (citing ’209, 19:22-24).) But Redfern also discloses and renders obvious textually verbatim “**duplicate queries**.” For example, Redfern identifies an Excite query of “Monarch+AND+(butterflies+OR+spend)+AND+(winter).” This is identical to the AltaVista query, “Monarch+AND+(butterflies+OR+spend)+AND+(winter).” (*Id.*)

Lastly, the Lycos, AltaVista and Excite search engines (as well as others Redfern describes (Redfern, 9:4-9, 9:66-10:1)) are clearly “**distinct information sources**”—the whole point of Redfern’s meta-search application is to enable a user to conveniently query multiple distinct information sources at once. (Redfern, 2:47-51, 2:37-40.)

(h) “asynchronously receiving...” (6[h])

Redfern describes “**asynchronously receiving multiple results to the multiple asynchronous queries**” from the search engines (“**information sources**”). (Redfern, 10:38-50.) Moreover, for substantially the same reasons as those discussed above for submitting “asynchronous” queries, “asynchronously receiving” the query results would have been obvious. (Schmandt, ¶226.)

(i) “scoring...” (6[i])

Redfern describes a process of determining the most relevant search hits to present to the user (i.e., “**determin[ing] one or more best responses to the at least one question**”) that involves “**scoring**” search hits (“**the multiple asynchronously received results**”) by matching words and phrases from the user’s natural language input against text from each search hit. (Redfern, 10:44-11:10.) The search hit(s) having the highest scoring “segments,” or excerpts, with scores “no less than a preselected level” are selected to present to the user. (Redfern, 13:9-11, 14:41-46, 15:19-31, 34:13-35 (highest scoring segments for question “Where do Monarch butterflies spend the winter?”).)

Redfern’s process, under the combination, “**us[es] a dynamic set of prior probabilities or fuzzy possibilities**” to determine the most relevant search hits because the process would have used the recognized words and phrases provided

from the user, which as discussed for claim 1 were determined using a dynamic set of prior probabilities in the trigram language model. (Schmandt, ¶229.)

As explained below, it would have been obvious under the combination for “scoring the relevance associated with the multiple asynchronously received results to determine the one or more best responses [to] include[]” the following steps.

(j) “determining...” (6[j])

Redfern discloses “**formulat[ing] a response to the at least one question**” by describing a process for identifying and extracting the best “segment,” or portion, of web pages corresponding to the search hits returned by each of the search engines that are responsive to the user’s question. (*See generally* Redfern, 10:45-15:41, Figs. 6-6c.) Redfern further discloses that this process involves “**determining one or more tokens**,” i.e., identifying words and phrases in the user’s search data that are used in scoring the search hits.

A “**token**” in information processing is generally understood as an individuated or non-reducible text element in data being parsed. (Schmandt, ¶232 (citing EX1009, p.0035 (“token”)).) Redfern explains that one of the first steps in identifying and extracting the best “segment”—that is, a portion of a response from a search engine that contains matches to groups of words or phrases in the search data corresponding to the user’s question—is to treat each group in the search data

as a separate candidate for scoring. Redfern describes example search data having four separate scoring candidates: (1) “or(phrase(contact), phrase(personnel), phrase(names))”; (2) “phrase(people)”; (3) “rank1(Gravis)”; and (4) “rank1(Logitech).” (Redfern, 10:64-11:10.) Redfern’s scoring candidates contain **“tokens required to formulate a response to the at least one question”** because they specify individuated words and phrases from the search data that are used to identify a segment and determine whether the segment meets a scoring threshold for being responsive to the user’s question. (Schmandt, ¶233.) In Redfern’s example above, the tokens would include “contact,” “personnel,” “names,” “people,” “Gravis,” and “Logitech.” In Redfern’s example question “Where do Monarch butterflies spend the winter?,” which is translated into the search data “rank1(Monarch), phrase(butterflies spend), phrase(winter),” the tokens would include “Monarch,” “butterflies spend” (and separately “butterflies” and “spend”) and “winter.” (Redfern, 11:4-30, 15:43-49.)

Redfern further explains that, if a score for a segment is too low, such as below a required minimum preselected level, then it is not provided as a response to the user. (Redfern, 15:19-31.) Therefore, a POSITA would have understood that the tokens were **“required to formulate a response to the at least one question”** because if the token match failed to reach the minimum level, then no segments

(“**response to the at least one question**”) would be provided to the user.

(Schmandt, ¶234.)

(k) “**applying...**” (6[k])

Redfern explains that search hits (“**multiple asynchronously received results**”) generally include a URL to the hit’s information source; e.g., in the hit shown below, the URL <<http://www.isit.com/butterfly/articles/sendmon.htm>>:

91% Monarch Population Plummets in Mexico [More Like This]  
URL: <http://www.isit.com/butterfly/articles/sendmon.htm>  
Summary: Urquhart tagged the first monarchs 59 years ago--and  
graciously agreed to tag the first paper monarchs for this symbolic  
migration! All migratory monarchs east of the Rocky Mountains spend  
the winter in just 9 major sanctuaries in Mexico!

(Redfern, 28:37-42 (Excite example); *see also id.*, 24:38-43 (AltaVista example).)

Redfern “**extract[s] one or more values for the asynchronously received results**” by having the meta-search application automatically download web pages pointed to by search hit URLs and search through the downloaded web pages to find and record values in the pages matching search data tokens (for example, finding and recording “Monarch” in the web pages). (Schmandt, ¶236; Redfern, 10:48-63 (“...retrieves the complete information sources **54** for each of a preselected maximum number of hits from each search engine **36**.”), 11:31-38 (“[T]he information source is examined to determine each match between its contents and the groups in the score table. For each match, an entry is made in the score table....”).)

Redfern further discloses that, to extract the one or more values, “**one or more scraping criteria**” are applied to the search hits (“**multiple asynchronously received results**”). “**Scraping**” in this context generally refers to techniques where a computer program extracts data from human-readable output from another program. (Schmandt, ¶237.) Redfern’s meta-search application extracts the value(s) by applying “**one or more scraping criteria**,” such as identifying and removing redundant search hits, retrieving a preselected maximum number of hits from each search engine, and then searching each retrieved hit for values matching the specific tokens corresponding to the search data. (*Id.*; Redfern, 10:48-63, 11:31-38.)

(I) “**evaluating...**” (6[I])

Redfern, under the combination, discloses “**evaluating the one or more extracted values using the dynamic set of prior probabilities or fuzzy possibilities**” because the words and phrases in the retrieved web pages are matched and scored using tokens resulting from the speech recognition process described in Coffman. (Schmandt, ¶238.) As explained, that speech recognition process uses a “dynamic set of prior probabilities” related to, for example, a trigram language model. For example, under the combination, the user’s question “Where do Monarch butterflies spend the winter?” would have been recognized from the user’s speech using prior probabilities of a language model. Words in the user’s question

are then used as search data tokens “Monarch,” “butterflies spend” (and separately “butterflies” and “spend”), and “winter.” (Redfern, 11:4-10, 15:43-49.)

Redfern’s process of matching and scoring (“**evaluating the one or more extracted values**”) includes “**resolv[ing] ambiguous, incomplete, or conflicting information associated with the one or more extracted values**” because it identifies the best-matching responses from the search hits and excludes results that are incomplete or ambiguous and score below a preselected level, such as segments that are shorter in length or poorer matches. (Schmandt, ¶239 (citing Redfern, 11:39-47 (exclude results having fewer than three matches), 13:1-14 (scoring calculation weights against segments that are shorter in length, and against hits with a high number of matches in just a few tokens and few matches for remaining tokens).)

**(m) “selecting...” (6[m])**

Redfern’s meta-search application “**select[s] one or more best values for the one or more required tokens from the one or more evaluated values**” by selecting the best segments including values matching the tokens as a result of the scoring process. (Schmandt, ¶¶240-241; Redfern, 15:19-28 (highest scoring segments “no less than a preselected level” selected for output to user).) For example, for the user’s question “Where do Monarch butterflies spend the winter?,” Redfern identifies the segment shown below as the best response, with highlighting added to indicate values matching tokens:

Ranking: 3.3608

From: dnr.state.il.us

Found with: Alta Vista

While most insects hibernate, the monarch is the only species of butterfly which actually flies to warmer weather (migrates) in winter. Monarchs from Illinois spend their winters in California and Mexico. In the fall, people have reported seeing entire trees covered with thousands of migrating monarchs! However, only about 1 percent of these monarchs actually survive the journey back to Illinois.

(Redfern, 34:15-23; *see also id.*, 24:38-43 (search hit corresponding to best segment), 16:44-45.)

**(n) “presenting...” (6[n])**

As explained for 1[f], Coffman discloses presenting to the user results generated by the conversationally-aware application (“**presenting the generated response to the utterance**”) “**via the speech unit,**” for example, by using text-to-speech to output results at telephone 600. (Coffman, Fig. 6, 40:9-19, 13:8-13, 58:9-11, 39:3-5.) Under the combination with Redfern, 6[n] simply corresponds to outputting the results of Redfern’s metasearch application based on a user’s request that included a command and question—for example, using text-to-speech to present the highest-scoring segments for the user’s search query “Where do Monarch butterflies spend the winter?” or “Search Lycos, AltaVista and Excite for where do Monarch butterflies spend the winter?” (Schmandt, ¶243.)

It would have been obvious to output the results of Redfern’s metasearch application “**via the speech unit,**” such as the speaker of telephone 600 as described

by Coffman. (Schmandt, ¶244.) Doing so would be consistent with Coffman’s goal of a universal conversational computing system useable across a wide range of user devices, including those with limited input/output capabilities beyond voice. (*Id.*; Coffman, 12:10-17, 9:12-22.)

## 2. Claim 9

It would have been obvious under the combination for **“scoring the relevance associated with the multiple asynchronously received results to determine the one or more best responses”** to further include the following steps:

### (a) “determining...” (9[a])

As explained, Redfern discloses only providing responses whose scores are no less than a preselected level (Redfern, 15:19-27), and the scoring process and use of a preselected level operates to resolve ambiguous, incomplete or conflicting information by eliminating search hits whose low scores would reflect those hits having ambiguous, incomplete or conflicting information. A POSITA would have therefore understood and found it obvious that a determination of no response would represent a determination that the matching and scoring process (use of **“the dynamic set of prior probabilities or fuzzy possibilities”**) **“failed to resolve the ambiguous, incomplete, or conflicting information associated with the one or more extracted values.”** (Schmandt, ¶246.)

(b) “requesting...” (9[b])

Under the combination, in a situation where the meta-search application failed to provide a response to the user’s question, it would have been obvious to prompt the user to refine/restate their question (“**requesting a user that spoke the utterance to provide additional information relating to the request**”) and resubmit to identify a search hit meeting the meta-search application’s scoring criteria (“**wherein the domain agent uses the additional information relating to the request to resolve the ambiguous, incomplete, or conflicting information associated with the one or more extracted values**”). (Schmandt, ¶247.) Identifying one or more search hits meeting the meta-search application’s scoring criteria would thus “**resolv[e] the ambiguous, incomplete, or conflicting information.**” (*Id.*) For example, Redfern describes “not obtaining acceptable results without multiple and/or recursive search attempts.” (Redfern, 2:9-11.) Coffman describes prompting the user for missing or ambiguous information. (Coffman, 41:16-21.) Applying Coffman’s known technique to improve Redfern’s similar process would have been obvious. (Schmandt, ¶248.)

## VII. CONCLUSION

Petitioner respectfully requests IPR institution.

Dated: April 5, 2024

Respectfully submitted,

COOLEY LLP  
ATTN: Patent Group  
1299 Pennsylvania Avenue NW  
Suite 700  
Washington, DC 20004  
Tel: (650) 843-5001  
Fax: (650) 849-7400

By: / Heidi L. Keefe /  
Heidi L. Keefe  
Reg. No. 40,673  
Counsel for Petitioner

## **CERTIFICATE OF COMPLIANCE WITH WORD COUNT**

Pursuant to 37 C.F.R. § 42.24(d), I certify that this petition complies with the type-volume limits of 37 C.F.R. § 42.24(a)(1)(i) because it contains 13,894 words, according to the word-processing system used to prepare this petition, excluding the parts of this petition that are exempted by 37 C.F.R. § 42.24(a) (including the table of contents, a table of authorities, mandatory notices, a certificate of service or this certificate word count, appendix of exhibits, and claim listings).

DATED: April 5, 2024

COOLEY LLP  
ATTN: Patent Docketing  
1299 Pennsylvania Avenue NW  
Suite 700  
Washington, D.C. 20004  
Tel: (650) 843-5001  
Fax: (650) 849-7400

/ Heidi L. Keefe /  
Heidi L. Keefe  
Reg. No. 40,673

## **CERTIFICATE OF SERVICE**

I hereby certify, pursuant to 37 C.F.R. Sections 42.6 and 42.105, that a complete copy of the attached **PETITION FOR INTER PARTES REVIEW OF U.S. PATENT NO. 8,015,006 B2**, including all exhibits (**Nos. 1001-1023**) and related documents, are being served via Federal Express on the 5th day of April, 2024, the same day as the filing of the above-identified document in the United States Patent and Trademark Office/Patent Trial and Appeal Board, upon Patent Owner by serving the correspondence address of record with the USPTO as follows:

174899 - The Law Offices of David A. Gerasimow, P.C.  
211 W. Wacker Dr.  
Ste. 1717  
Chicago, IL 60606

And, via Federal Express upon counsel of record for Patent Owner in the litigation pending before the U.S. District Court for the District of Delaware entitled *Dialect, LLC v. Google LLC*, Case No. 1:23-cv-00378 (D. Del.) as follows:

Garland Stephens  
BLUE PEAK LAW GROUP LLP  
3139 W Holcombe Blvd, PMB 8160  
Houston, TX 77025  
Telephone: 281-972-3036  
garland@bluepeak.law

DATED: April 5, 2024

/ Heidi L. Keefe /  
Heidi L. Keefe  
Reg. No. 40,673  
COOLEY LLP  
1299 Pennsylvania Ave. NW  
Suite 700  
Washington, D.C. 20004  
Tel: (650) 843-5001